Comparative analysis of EHR-based stroke phenotyping methods, their applications, and interpretation

Phyllis Thangaraj, Benjamin Kummer, Tal Lorberbaum, Mitchell Elkind, Nicholas Tatonetti

Columbia University Department of Biomedical Informatics, New York, NY, United States

Stroke is the second leading cause of death in the world and top cause of disability in the US. The plurality of electronic health records (EHR) provides an opportunity to study this disease *in situ*. Doing so requires accurately identifying stroke patients from medical records. So-called "EHR phenotyping" algorithms, however, are difficult and time consuming to create and often must rely on incomplete information. There is an opportunity to use machine learning to speed up and ease the process of cohort and feature identification. We systematically compared and evaluated the ability of several machine learning algorithms to automatically phenotype acute ischemic stroke patients. We found that these algorithms can achieve high performance (e.g. average AUROC=0.955%) with little to no manual feature curation, and other performance evaluators differentiate each model's ability to generalize. We also found that commonly available data such as diagnosis codes can be used as noisy proxies for training when a reference panel of stroke patients is unavailable. Additionally, we find some limitations when the algorithms are used to place patients into stroke risk classes. We used these models to identify unidentified stroke patients from our patient population of 6.4 million and find expected rates of stroke across the population.

Keywords: phenotyping algorithms, ischemic stroke, electronic health record studies

1. Introduction

Stroke is a highly heterogeneous and complex disease that is a leading cause of death and severe disability for millions of survivors worldwide.[1] It is characterized by an acute focal loss of neurological function and is primarily caused by loss of blood flow to a specific area of the brain. There are many identifiable risk factors for stroke, which include various metabolic, cardiovascular, and coagulative diseases, medications, lifestyle, and demographics. Triggers

such as pollution, infection, and inflammatory disorders, further complicate the etiology of the disease.[2] Most of the unidentified risk, up to 40%, is thought to be genetic.[3] Accurate determination of the etiology of disease is essential for risk stratification and optimal treatment, but this can be difficult as up to 35% of strokes are of undetermined cause.[4,5]

Traditionally, identifying a stroke patient requires the integration of multiple facets of data including medical notes, labs, imaging reports, and medical experience by neurologists. This requires time consuming manual review. Stroke diagnoses have also been missed or falsely assigned[6] . Stroke is often coded in outpatient follow up, so a hospital EHR may not have the ICD9 or 10 data to identify stroke with high sensitivity. Given the incompleteness of identifying patients using stroke-specific ICD9 codes and the availability of structured multimodal data in EHR settings, there is a need to move beyond laborious manual chart review and to automate identification of stroke patients with commonly accessible EHR data.

Phenotyping algorithms must address two tasks: curating features to define the phenotype, and identifying cases and controls. Computable and reproducible phenotypes should optimally include rule-based algorithms, structured data, and standardized terminologies.[7] Possible features from the electronic health record may include insurance billing diagnosis and procedure codes, medications, laboratory values, imaging, and terms parsed from natural language processing of medical notes.[8] While many model feature are curated by physicians, Often, the features chosen for modeling are curated by physicians, naïve feature sets have appropriate machine learning applications.[9–11] Recent methods also have combined publicly available sources such as Wikipedia and Medscape with EMR data to automatically curate features.[12,13] The second task of cohort curation in EHR phenotyping, can also be automated through machine learning.[10,14,15] Most methods have relied on a small number of expert curated cohorts, though recent studies have applied noisy labeling or automated feature selection without manual curation.[16–18]

Another important trade off includes application of feature reduction. Structured medical

data, in particular, have hierarchical organization that can be utilized for the grouping together of similar features and have improved classification performance.[19,20]

Recent phenotyping algorithms for stroke have used machine learning to improve classification performance and expand cohort size.[15,21–23] Additionally, recent work has shown that machine learning classifiers using large clinical feature sets may identify cohorts with AIS with superior performance than existing classifiers using diagnosis codes. Both algorithms successfully use the Tirschwell and Longstreth (T-L)-defined acute ischemic stroke codes within their algorithms to detect patients with stroke over diagnosis code identification and use a gold standard of physician labeled cases and controls. In contrast to these approaches, we attempt to recover patients with an acute ischemic stroke were were not given the diagnosis code. We also determine the feasibility of training a phenotyping algorithm with a code-based cohort versus a silver-standard curated cohort. We also ascertain the clinical utility of the model output by calibrating the models.

We compare and evaluate the ability of 75 machine learning models to identify patients with acute ischemic stroke. Given the inconsistency of diagnosis code implementation and the difficulty of identifying true controls in the EHR, we curate and compare multiple case and control sets to determine the best combination for phenotyping. We also apply key methods to optimize the robustness of these models to missing data, the calibration to ensure a clinically meaningful model output, and the number of features to improve generalizability. Finally, we classify patients previously unidentified in the EHR as potential stroke patients that do not have ICD9 or ICD10 codes for acute ischemic stroke. Our phenotyping method utilizes machine learning classifiers to increase the number of stroke patients recovered within the EHR and reduce the time and effort needed to find them for research studies.

2. Results

2.1. *Data Sources*

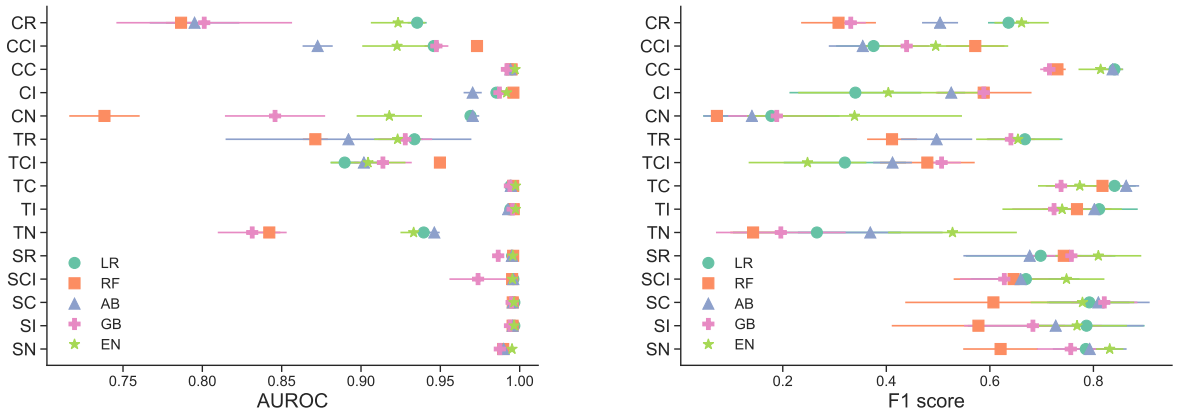
The CUIMC Clinical Data Warehouse (CDW) contains longitudinal health records of 6.4 million patients spanning from 1985-2018. Of these patients, 4,844 were stroke service patients and served as our primary positive controls (i.e. cases). We (PT and BK) each manually reviewed 50 of these patients' charts and determined a false positive rate of 4 and 16% respectively. We defined an additional two case sets (1) to explore the performance when using alternative training sets.

Variable	Identification	N Samples
Total Patients	CUIMC CDW Person ID	6,377,222
Diagnosis Codes	ICD9, ICD10, SNOMED	140,300,457
Procedure Codes	ICD9, ICD10, CPT, SNOMED	64,373,775
Prescription Orders	RxNorm	78,959,941
Training Categories		
Stroke Service Cases (S)	Seen by NYP Stroke Service	4,844
Tirschwell Criteria AIS Cases (T)	ICD9: 434.x1, 433.x1, ICD10: I63.xxx	79,306
CCS Cerebrovascular Cases (C)	ICD9:346.6x,430,431,432.x,433.xx	174,248
AIS Mimetic Diseases Controls (N)	ICD9:191.x,225.x,340,250.0,431	8,438
Without Stroke Controls (I)	No (T) Codes	5,243,646
Without Cerebrovascular Disease Controls (C)	No (C) Codes	5,148,703
With Cerebrovascular disease, w/o AIS Controls (CI)	(T) codes, No (C) codes	94,985
Random set of patients Controls (R)	At least one ICD9 or ICD10 diagnosis code	5,396,172

Table 1: Select Structured Data and Sample Case/Controls for models available in CUIMC Common Data Warehouse.

2.2. *Performance of Algorithms*

After training the machine learning (ML) algorithms (see methods 4.1), we calculated the area under the receiver operating curve and average precision score, which is a proxy for area under the precision-recall curve, for each of the 90 classifier type and case-control combination of models (3 case sets x 5 controls sets x 6 ML models) on a hold-out test set of 1,000 stroke service patients and 100,000 random patients in the electronic health record. Logistic regression classifiers with L1 penalty gave the best AUROC performance (0.89-0.997) and the best average precision score (0.685-0.969), followed by logistic regression classifiers with elastic



(a) AUROC

(b) F1 score

Fig. 1: Performance of select models on holdout test set (left: AUROC, right: F1). (LR) logistic regression with l1 penalty, (RF) random forest, (AB) AdaBoost, (GB) gradient boosting, (EN) logistic regression with elastic net penalty. Different combinations of cases and controls are shown on the y-axis. Cases may be one of cerebrovascular (C), T-L (T), or Stroke Service (S), see methods 4 for definition of case sets. Controls may be one of random (R), cerebrovascular disease but no AIS code (CI), no cerebrovascular disease (C), no AIS code (I), or a stroke mimetic disease (N), see methods 4 for definition of control sets. Threshold to compute the F1 score on the testing set was chosen as the threshold that yielded the maximum F1 in cross-validation on the training set, see methods 4.1, table 6.

net penalty (Figures 1a, Table 2). Neural network models (N=15) gave low performance and were not included in the rest of the analysis.

Across all classifier types, models using stroke service patients as cases had the best F1 scores, while the other cases with absence of cerebrovascular disease as controls had F1 scores above 0.8 for all classifier types (Figure 1b, Table 6). Stroke mimetic disease controls fared the worst in performance. The stroke service cases gave the best performance overall, however, the cases identified through AIS codes defined by Tirschwell and Longstreth (T-L) and controls without cerebrovascular disease or the AIS codes give high precision as well. The sensitivity of the models ranged widely, between 0.18 and 0.95, while specificity narrowly ranged between 0.86-1.0 (Table 3).

Case/ Control Combo	LR AUROC	RF AUROC	AB AUROC	GB AUROC	EN AUROC	LR Ave. Precision	RF Ave. Precision	AB Ave. Precision	GB Ave. Precision	EN Ave. Precision
SN	0.989	0.99	0.99	0.987	0.995	0.916	0.81	0.902	0.891	0.929
SI	0.997	0.996	0.996	0.994	0.996	0.952	0.886	0.94	0.95	0.945
SC	0.997	0.996	0.996	0.995	0.996	0.946	0.869	0.957	0.958	0.94
SCI	0.996	0.995	0.996	0.974	0.995	0.969	0.957	0.963	0.837	0.963
SR	0.995	0.996	0.995	0.986	0.996	0.966	0.924	0.964	0.929	0.97
TN	0.94	0.842	0.946	0.831	0.933	0.748	0.085	0.686	0.372	0.698
TI	0.994	0.996	0.993	0.996	0.997	0.928	0.911	0.884	0.846	0.943
TC	0.994	0.996	0.994	0.994	0.997	0.935	0.923	0.925	0.828	0.944
TCI	0.89	0.95	0.902	0.914	0.904	0.752	0.737	0.407	0.549	0.781
TR	0.934	0.871	0.892	0.928	0.923	0.748	0.401	0.678	0.705	0.756
CN	0.969	0.738	0.97	0.846	0.918	0.769	0.031	0.526	0.214	0.535
CI	0.985	0.996	0.97	0.987	0.992	0.885	0.916	0.582	0.639	0.905
CC	0.995	0.995	0.994	0.992	0.997	0.919	0.821	0.914	0.816	0.931
CCI	0.946	0.973	0.873	0.947	0.923	0.842	0.813	0.65	0.537	0.798
CR	0.935	0.787	0.795	0.801	0.923	0.685	0.259	0.413	0.36	0.692

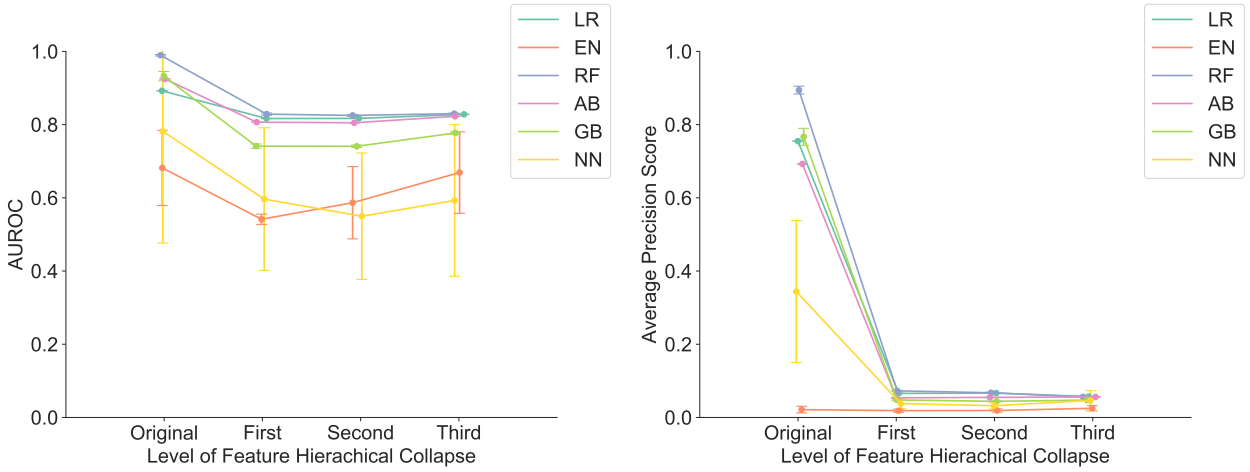
Table 2: Area Under the Receiver-Operating Characteristic Curve and Average Precision score across models for Hold-out Test set. Blue values > 0.9, yellow values between 0.7 and 0.9, red values < 0.7.

2.3. *Feature Importances and Score Generation*

We aggregated the most common features found in the top ten in importance or beta coefficient weight for each model, as seen in figure 4. We found the most commonly chosen features were procedures used in evaluation of AIS, including extra- and intra-cranial arterial scans, CT scans and MRIs of the brain, and MR angiography. We also found age over 50 to be a top common demographic feature, and Aspirin and atorvastatin, a cholesterol lowering medication, to be important medication features as well. Top diagnosis features included history of TIA or stroke w/o residual effect, hemiplegia, affected unspecified side, intercerebral hemorrhage, convulsions, and iatrogenic cerebrovascular infarction or hemorrhage.

2.4. *Feature Reduction and subsequent algorithm performance*

The number of features of the models ranged from 22,000-35,500 depending on the samples. We evaluated the performance of collapsing features by ICD9/ICD10 and Anatomical Therapeutic Chemical Classification System (ATC) code hierarchy, reducing the number of features per model to 72-1500 features, depending on the class size and collapsing level (see methods 4.4).



(a) AUROC stratified by classifier type (b) Average Precision Score stratified by classifier type

Fig. 2: Performance of models at increasing level of feature hierarchical collapse. Logistic regression with L1 penalty (LR), Logistic Regression with elastic net penalty (EN), random forest (RF), AdaBoost (AB), gradient boosting (GB), and a two layer neural network (NN). (a) The area under the receiver operating characteristic curve (AUROC) for each of the models versus the level of abstract with 'First' being the highest level of abstraction and 'Third' being the least. (b) Average precision, annotated described in (a).

As seen in 2, we found a 9-27% reduction in AUROC performance, compared to a 90-94% reduction in average precision score.

2.5. *Calibration*

We calibrated the models to produce meaningful probability estimates since they were trained with approximately a 1:1 ratio of cases and controls and, in general, we expect a 1:100 ratio of stroke patients to all other patients in our EHR (methods 4.1), figures 3 and supplementary material). We found a wide range of success in calibration of the models. Uncalibrated, the AdaBoost classifier produced a narrow range of probabilities, between 0.4 and 0.6, while the other models ran a range of probabilities from 0 to 1. On average, test set calibration of the AdaBoost models had the lowest root mean squared error (RMSE). The AdaBoost models with controls without cerebrovascular disease codes (C) showed the lowest rmse for each respective case set (rmse=0.047 (SC), 0.1024 (TC), and 0.0637 (CC)). AdaBoost and Logistic

regression models with stroke service (S) cases and stroke mimetic controls (N) also showed low rmse (0.0782 and 0.1177). Models with no calibrated predicted probabilities above 0.1 were given an rmse of N/A, and this applied to 18 of the models. Most of them used a case set of patients with any cerebrovascular disease code (C) or used a gradient boosting classifer. Notably, the logistic regression models (L1 and EN penalty) and gradient boosting models with S cases and random (R) controls or controls with cerebrovascular disease but no stroke (CI) had uncalibrated test set probabilities with low rmse, but high rmse after calibration (see figure 12, figure 11).

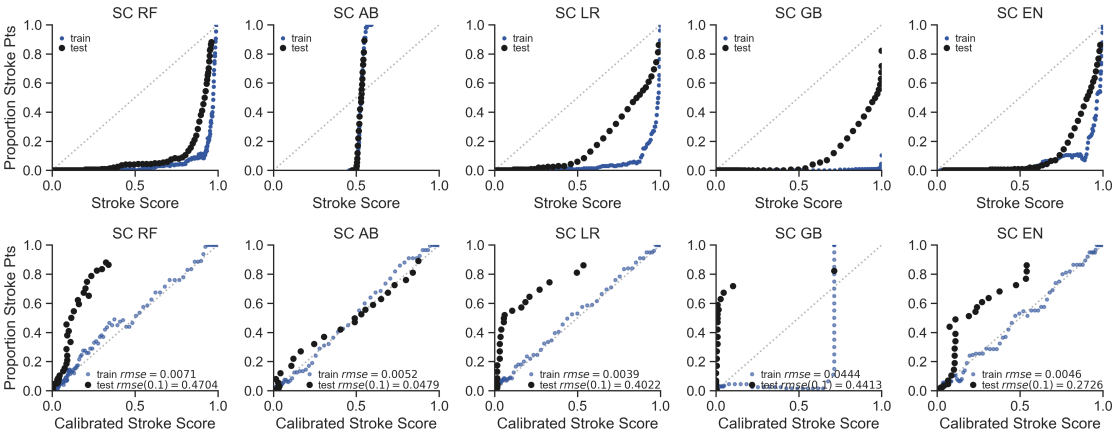


Fig. 3: Classifier type with SC case-control combination varies in calibration success between stroke score and actual proportion of patients at each probability. Top panel plots the proportion of stroke case patients within a bin of 100 patients with similar stroke scores for models trained with stroke service cases and controls without cerebrovascular disease icd9 or 10 codes. The black circles plot the mean test stroke score across all training set folds, while the blue circles plot the training fold stroke scores combined. Classifier type varies from left to right:Random Forest (RF), AdaBoost (AB), Logistic Regression with L1 penalty (LR), Gradient Boosting (GB), Logistic regression with elastic net penalty (EN). Bottom panel plots the proportion of stroke patients within a bin of 100 test set patients with similar stroke scores versus the calibrated test scores at specific scores rounded to the nearest thousandth (Black dots). The test scores are calibrated from an empirical distribution determined from the training set 4.3. The blue dots plot the proportion of stroke patients within a bin of 100 bootstrapped training set patients with similar stroke scores versus the calibrated training set scores rounded to the nearest thousandth. The light grey dots show perfect calibration.

2.6. *Robustness*

We evaluated the robustness of the models to increasingly smaller training sets (Figure 4). When stratified by classifier type, all classifier types except AdaBoost maintained AUROC above 0.9 up to 5% of the training size (350-800 samples) , and maintained an AUPRC above 0.9 up to 10% of the training size (700-1600 samples).When stratified by case, models with stroke service cases (S) maintained AUROC above 0.9 up to 1% of the training size (70-160 samples) and up to 5% of the training size. Other cases (T,C) maintained AUROC and AUPR above 0.9 up to 30% of the training size (2100-4800 samples). When stratified by control, models with stroke mimetic controls (N) maintained AUROC and AUPRC above 0.9 at up to 10% of the training size. In (4), the points below 1% of the training size were removed from the stroke mimetic control sets because of very high error from very small training sizes between 6-30 patients. Random patient controls (R) maintained AUROC and AUPRC above 0.9 at up to 5% and 30% of the training size respectively. Controls without T-L AIS codes (I) maintained AUROC and AUPRC above 0.9 at up to 1% and 5% of the training size respectively, while controls without cerebrovascular disease codes (C) maintained AUROC and AUPRC above 0.9 at up to 0.5% (35-80 samples) and 5% of the training size respectively. Finally, controls with cerebrovascular disease but without T-L AIS codes maintained AUROC and AUPRC above 0.9 at up to 10% and 30% of the training size respectively.

2.7. *Feature Category Contribution to Models*

We trained models on individual feature type categories (procedures, medications, conditions, and T-L AIS conditions) and compared their contributions to model fit using nested linear models (Methods 4.5). We found that procedures contributed the most to model probabilities (Figure 5a). 95% of the generalized linear model (GLM) models had a significant ($p < 0.05$) coefficient weight for the procedure feature category, followed by medication orders, which made a significant contribution in 60% of the models. The T-L AIS codes feature category,

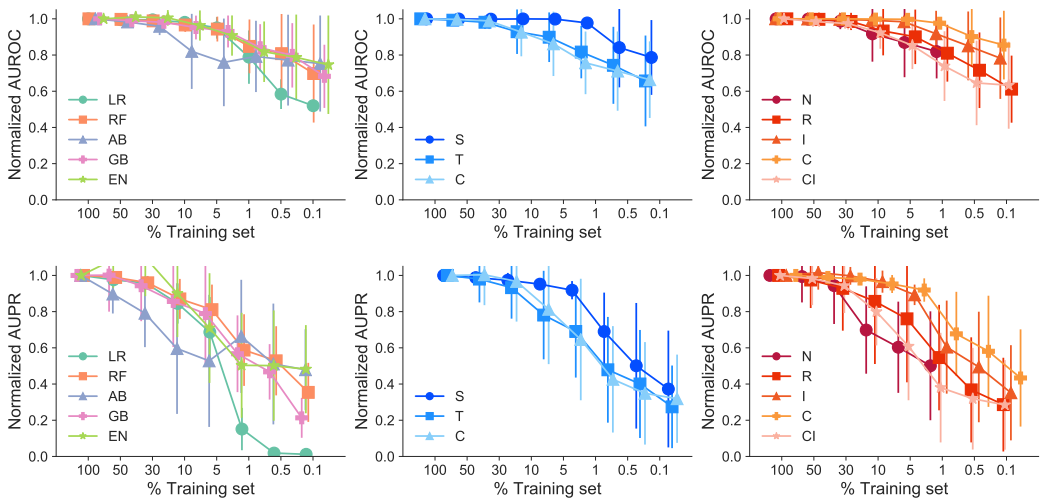
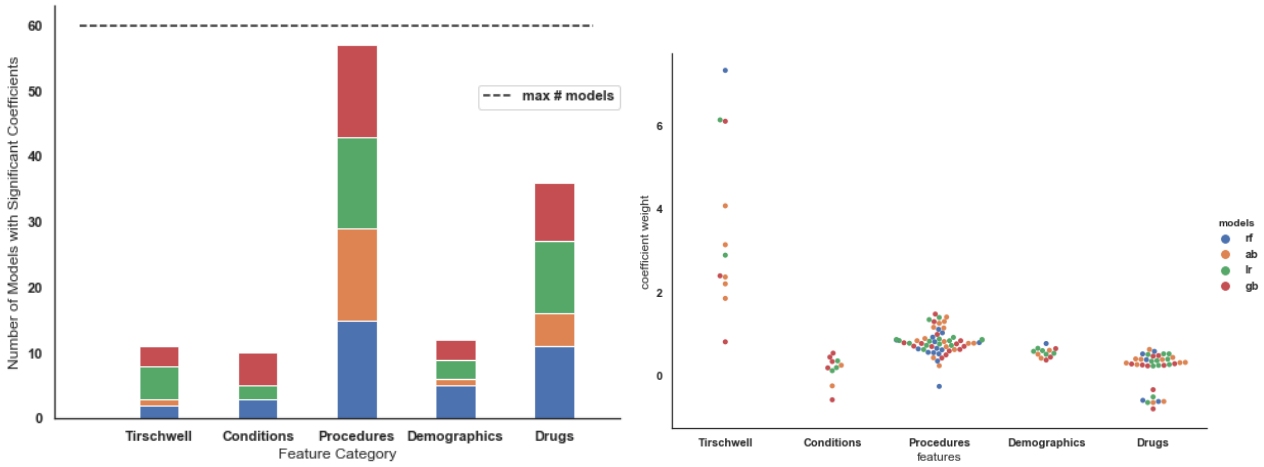


Fig. 4: Robustness of Models. Top panel, Left to Right: Normalized Area under the receiver operating curve (AUROC) across all case-control combinations, stratified by classifier type (LR, RF, AB, GB, and EN), normalized AUROC across all models stratified by case type (S, T, C), normalized AUROC across all models stratified by control type (N,R,I,C,CI). Bottom Panel, left to right: normalized Area under the Precision-Recall Cure (AUPR) values across all case-control combinations, stratified by model, normalized AUPR across all models stratified by case type, normalized AUPR across all models stratified by control type.

notably, was significant in 12% of the models. The beta coefficient of the T-L category was 2-8 times larger than the other category coefficients.

2.8. *Additional Patients predicted from EHR*

We applied our models to the entire NYP EHR to determine whether there are AIS patients without a ICD9 or ICD10 codes defined by T-L (methods 4.6). We trained models using our 15 case-control combinations and 5 classifier types and then applied the models to the entire CUIMC EHR with at least one icd9 or 10 diagnosis code, totaling 5,324,725-5,315,923 patients depending on the case/control set. Based on the thresholds defined by the maximum F1 score from the training set (see methods 4.1), we determined the prevalence of AIS patients estimated in the EHR by each model. We found that the results varied widely across models, but most predicted that a prevalence between 0.2-2% of patients in the EHR was an acute



(a) Significant feature categories across models (b) Coefficient weights for feature categories

Fig. 5: Feature category contribution to model fits. (a) shows the total number of models with a significant coefficient weight for each feature category, stratified by classifier type. (b) shows the value of the significant coefficient weights, stratified by classifier type.

ischemic stroke patient. The least sensitive models (see table 3), which included patients with stroke mimetics as controls and logistic regression, AdaBoost, and gradient boosting classifiers predicted a very high prevalence of above 97% (see table 3). From these proposed cases, we also calculated the proportion of patients with a ICD9 or 10 code for acute ischemic stroke (AIS) defined by T-L. The most well calibrated models, (AdaBoost SC, TC, CC, TN) (see figure 3, figure 15, figure 20,figure 13) had sensitivities between 0.847-0.89, specificities 0.996-0.998, and estimated AIS prevelance in the EHR between 1.3-2.6% (see table 3,table 5). Within these proposed AIS patients, 22.4-45% of them had an AIS diagnosis code (see table 5).

3. Discussion

3.1. *Machine learned models are able to identify acute ischemic stroke patients without direct evidence.*

Using a feature agnostic, data-driven approach, we have developed models that identify acute ischemic stroke (AIS) patients from the EHR without AIS ICD9 codes defined by Tirschwell and Longstreth (T-L), which are often incomplete or inaccurate. Unlike other methods, we

did not use these codes within our model features (see methods 4. Past machine learning phenotyping algorithms relied on manually curated features or have stroke codes (ICD9 430-438) as the sole nonzero features in their models.[15,21,24] We show that AIS patients can be recovered from other EHR-available structured clinical features excluding the T-L criteria.

3.2. *Case and control choices are important for acute ischemic stroke phenotyping.*

Identifying cases and controls for phenotyping algorithms can be a laborious process and difficult to define given the richness of the data available. Within the sparse, structured EHR diagnosis codes, patients without a stroke diagnosis code cannot always be considered a control for stroke cohort generation. Similarly, it is difficult to determine whether patients with cerebrovascular disease or metabolic disease, which notably have overlapping genetic etiologies with acute ischemic stroke, should be considered controls as well. Neurological events from diseases such as multiple sclerosis or hypoglycemia may be mistaken for an acute ischemic stroke, which can further muddy cohorts defined from ICD9 or ICD10 diagnosis codes. We systematically analyzed our 15 difference case-control training sets, each of which addressed the ambiguities discussed above and found marked differences in performance. The stroke service cases (S) identified the stroke patients the most accurately and with the highest precision and recall. Pairing cases identified from AIS codes with controls from either 1) no cerebrovascular disease or 2) AIS codes give high precision as well (2). This suggests that a manually curated cohort may not be necessary to train the phenotyping models, and the T-L AIS codes may be enough to define a training set. That said, using these models, we were able to increase our stroke patient cohort by 60% across the EHR who do not have T-L AIS codes, meaning that the T-L AIS codes themselves are not sufficient.

3.3. *Procedures serve as a proxy for acute ischemic stroke diagnosis codes in model features.*

Our findings suggest that stroke evaluation procedures, such as a CT scan or MRI, are important features in most of the models and may serve as a proxy for the T-L AIS diagnosis codes. We compared the weight of each feature category of T-L AIS diagnosis codes, diagnosis codes other than T-L AIS codes, procedures, drugs, and demographics for classification by implementing a generalized linear model of the test set probabilities derived from each category. We found that procedures provided a significant contribution ($p < .05$) to patient classification in 95% of the models , while the T-L diagnosis codes provided a significant contribution in 20% of the models, though the T-L beta coefficient was 2-8 times higher than the other categories. This suggests that procedures are important proxy features for the T-L AIS diagnosis codes.

3.4. *Other diagnosis codes may be useful for phenotyping acute ischemic stroke.*

In addition, iatrogenic cerebrovascular infarction or hemorrhage (SNOMED ID 44834124, ICD9 997.02) was found to be a top feature, and is a code not within the cerebrovascular infarction ICD9 code class. Although this code does not distinguish between an ischemic and hemorrhagic stroke, it was found in 4% of cases and could be a useful addition for stroke cohort identification purposes. Other conditions that were highlighted by the models include hemiplegia and convulsions. This is consistent with features that may be coded during initial inpatient evaluation for a stroke. In some cases, the actual AIS will only be coded during outpatient followup. For example, within the holdout test set, 13.5% of the cases did not have AIS codes in the inpatient setting but did have in outpatient data. In addition, 90% of these patients had a head CT scan, supporting the potential recovery of these patients when including procedures in the feature set.

3.5. *Models showed robustness reduction of training set size, but not with code-hierarchy-based feature reduction.*

Performance of the random forest, logistic regression with L1 and elastic net penalties, and gradient boosting models was robust in terms of area under the receiver operating curve and average precision, even when up to 90% of the training set was removed. However, reducing feature set size by collapsing with CCS and ATC hierarchy resulted in large drops in performance, most likely due to the 95-99% drop in feature set size and high level description of the features. This suggests that the feature reduction by hierarchical collapse trains with excessively high-level features for successful AIS identification. Feature size reduction of up to 60% through training set size reduction, however, maintained high performance of the models.

3.6. *Calibration using an empirical function differentiates the models and may identify additional control sets*

We calibrated the output of the models so that the predicted probability could be meaningfully interpreted (see methods 4.3). The AdaBoost classifier type, in particular, produced calibrations that had the lowest root mean squared error (RMSE) of probabilities predicted. AdaBoost models that were trained with controls without cerebrovascular disease produced the lowest RMSE. This suggests the need for a more stringent definition of controls than one simply no stroke code diagnosis. AdaBoost and logistic regression models trained on stroke service patients as cases and patients stroke mimics as controls also performed well and showcases another important control set to consider. Some case-control sets, namely stroke service cases with random patients as controls and cerebrovascular disease without stroke codes as controls produced well-calibrated test sets before calibration by the boot-strapped training set but poor calibration in testing. This suggests that our training set may have over-weighted the importance of the controls. In addition, we evaluate the calibration performance based on the RMSE of the calibrated test set, when in the future we would like to evaluate based on the

training set. Our study shows that calibration potential of each model may better discriminate model success for studies with a large control:case ratio than traditional evaluation methods such as area under the receiver operating curve.

3.7. *Models can identify a large number of stroke patients without acute ischemic stroke diagnosis codes.*

Our results from traditional model performance and calibration evaluation show that our best machine learning phenotyping algorithm used an AdaBoost classifier trained with controls without any ICD9 or ICD10 cerebrovascular disease codes and a manually curated case cohort from a stroke service. A similar model, however, using cases identified by acute ischemic stroke diagnosis codes performed comparably well, suggesting these models do not require manual curation for high performance. With these models, we have shown that we can add approximately 60% more patients to our current T-L criteria-derived cohort by applying these models to the entire EHR.

3.8. *Limitations*

The study has several limitations. First, we relied on noisy labels and proxies for training our models, as evidenced by the false positive rate of 4-16% that was determined by manual review. Without a gold standard set of cases, model performance is difficult to definitively evaluate. However, we plan to validate with a physician derived set in the future. To make the model simple, reproducible, and data-agnostic, we also did not take into account the patient's timeline, used only structured data from specific clinical categories, and did not use clinical notes. Even given these limitations, we were able to identify the hold out test set with high precision. Further, we do not exhaustively explore all possible sets of cases and controls due to limitations of time and computational complexity.

3.9. *Conclusions and Future Directions*

We developed and compared several different machine learning classifiers to identify patients with acute ischemic stroke. This tool can be valuable for research purposes identifying patients for genetic studies or developing prediction models for stroke. We show that structured data may be sufficiently accurate for classification, allowing for widespread usability of the algorithm. In addition, patient cohorts derived using AIS diagnosis codes may obviate the need for manually-curated cohorts of patients with AIS, and procedure codes may be useful in identifying patients with AIS that may not have been coded with AIS-related diagnosis codes. Our study has several strengths over previous methods. Recent studies have either relied on manually curated features sets or boil down to essentially only using the T-L critera.[21] The former will be untenable for any large-scale analysis and the latter will have poor sensitivity.[15,21] By removing the T-L criteria from our model training and allowing any code to be used as a feature, we identify patients who may not have been immediately coded with acute ischemic stroke during inpatient treatment. We also have developed classifiers that phenotype patients with a probability of having had an acute ischemic stroke, which moves beyond binary classification of patients. The future of phenotyping calls for moving beyond binary classification and developing a more granular description of patient's disease state.[25–29] We plan to apply probability scores to expand cohort size of heritability and GWAS studies. We hypothesize that expanding cohort size by assigning a probability of disease will improve the power of these studies.

4. Methods

From the CDW, we gathered diagnostic and procedure insurance billing codes, medication prescriptions, race, ethnicity, age, and gender of each patient and classified each as a binary feature based on presence or absence of the feature. We mapped ICD9, ICD10, and CPT4 codes for diagnoses and procedures to SNOMED concept IDs, and also included SNOMED concept

IDs for medication prescriptions as features. SNOMED IDs have been shown to perform similarly to ICD9 and ICD10 codes for phenotyping.[30] Using an algorithm combining race and ethnicity codes, we identified patients with Hispanic ethnicity.[31] Most recent diagnosis in the medical record served as the age end point. To discretize age to a binary variable, we set ages greater than or equal to 50 to 1 and the remaining to 0. Importantly, we excluded codes used to curate the cases from our feature set. Within the case-control training sets, the cases include a sample of patients evaluated by Columbia's stroke service for acute stroke from the past 7 years curated by Columbia neurologists. Other case sets include patients in the EHR with the Tirschwell-Longstreth (T-L) criteria for acute ischemic stroke and associated ICD10 codes, and patients with any diagnosis code for cerebrovascular disease defined by the ICD-9-CM Clinical Classifications Software tool (CCS). The T-L criteria defines AIS patients with an ICD9 code of 434.x1, 433.x1, 436 (x = any number).[24] The ICD10 equivalent of these codes include I63.xxx and I67.89 based on CMS General Equivalence Mappings.[32] Controls include patients without CCS cerebrovascular disease codes, patients without T-L diagnosis codes for stroke, patients with diagnosis codes for CCS cerebrovascular disease other than ischemic stroke, neurology patients with diagnosis codes for stroke mimetic diseases such as hemiplegic migraine (346.3), brain tumor(191.xx,225.0), multiple sclerosis (340), cerebral hemorrhage (431), and hypoglycemia with coma (251.0), and a random sample of patients. Since there are over 5 million patients without a cerebrovascular diagnosis and a thus large imbalance in case and control samples, the controls were randomly sampled to create a 1:1 ratio of cases to controls. In addition, we set the maximum sample size to 16,000 to control feature size.

4.1. *Model Development*

We trained logistic regression classifiers with l1 and elastic net regularization, random forest, AdaBoost, gradient boosting, and neural network classifiers on the features gathered for fifteen combinations of cases and controls using the python scientific computing package sklearn. We

ran a grid search for the hyperparameters of maximum tree depth, number of estimators, l1-l2 ratio, penalty parameters, subsampling rate, learning rate, momentum, and dropout. Outside of the default parameters, we used a max tree depth of 100, 1000 estimators, and square root maximum feature number for the random forest models. For L1 Logistic regression, we used an inverse of regularization strength of 0.1, and for elastic net regularization we used a penalty parameter of 0.01 and l1-l2 ratio parameter of .01, and "log" loss. For the boosting algorithms, we used a learning rate of 0.1, number of estimators 1000, and for gradient boosting we also subsampled at a rate of 0.5, max depth as 10, and square root maximum feature number. The neural network model was comprised of two layers, the first with 64 neurons, relu activator, and l1 kernel regularizer. The second layer contained two neurons and a softmax activator. Learning was compiled by stochastic gradient descent with learning rate of 0.01 and momentum 0.9, nesterov=True. We included a dropout of data in the first node at a rate of 0.3. Loss was calculated by categorical cross-entropy. We evaluated the models using 10-fold cross validation. Evaluators were scored by the area under the receiver operating characteristic curve, sensitivity, and specificity. Within the testing set of each fold, controls were bootstrapped to form a 100:1 control to case ratio to represent the prevalence of acute ischemic stroke in the population. Precision and recall were then calculated from the bootstrapped set. To determine the optimal threshold to maximize precision while maintaining a high recall, we calculated maximum F scores at different βs using the following equation:

$$F_\beta = (1 + \beta^2) * \frac{precision * recall}{(\beta^2 * precision) + recall}$$

where β = 1.0,0.5,0.25, and 0.125. We also evaluated a test set using the same metrics on a held out set of 1000 patients from the stroke service and 100,000 non-overlapping random patients. Sensitivity, specificity, precision, and recall were evaluated by using the probability thresholds determined from cross-validation.

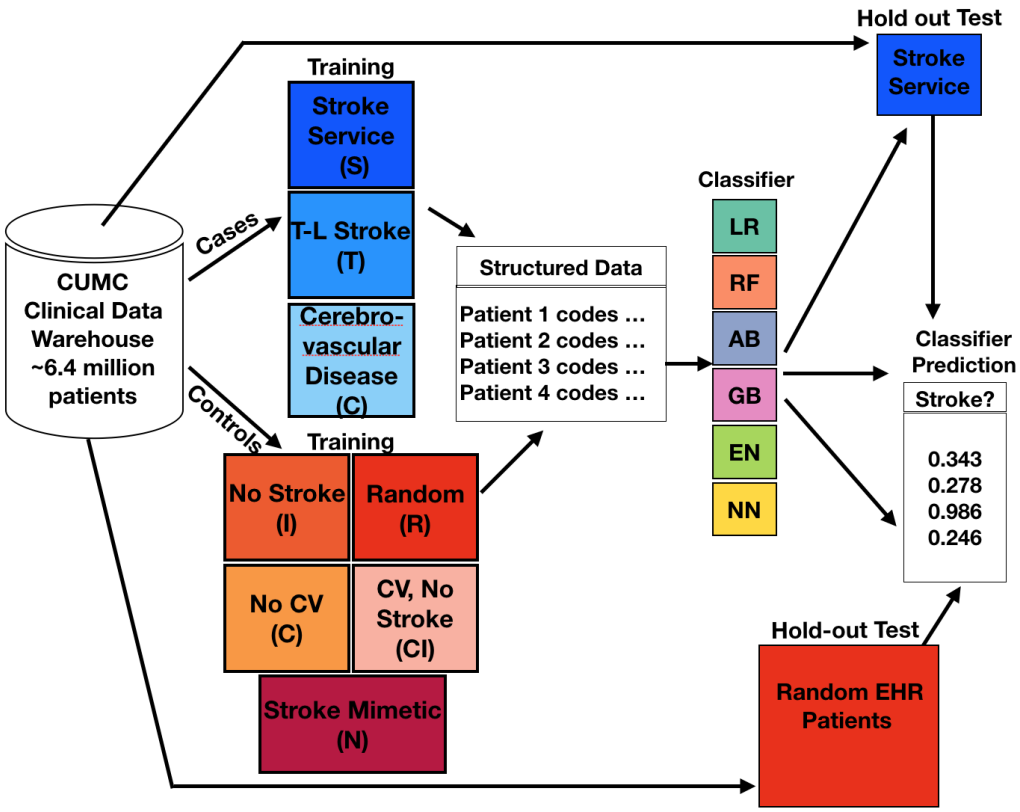


Fig. 6: Schematic of Model Training and Testing

4.2. *Robustness of Model*

To test the robustness of the models, we trained on 100%, 90%, 70%, 50%, 30%, 10%, 5%, 1%, 0.5%, and 0.1% of the original training set. We then evaluated the performance of each training size on the test set through area under the receiver operating curve (AUROC) and average precision score, which is a proxy for area under the precision recall curve (AUPR). We resampled the set 10 times at each training size and averaged the results.

4.3. *Calibration*

The classifiers assign a probability of being an acute ischemic stroke case to each patient. In order to interpret the meaning of these model probabilities and its applicability to other patients, we evaluated the calibration of the models. For each training fold during 10-fold cross validation, we first sorted the probabilities by value. We then averaged a bin of 100 patients

intervals based on their assigned probability (predicted probability). Within each bin, we calculated the proportion of cases (actual probability). We repeated this sliding bin averaging for every 10 patients, creating an empirical function between the training set predicted probabilities and actual probabilities averaged across training folds. We then repeated the same averaging of predicted and actual probabilities for the sorted test set. We then calibrated the test set predicted probabilities using the empirical function generated from the training folds. To evaluate calibration success, particularly for identifying cases, we calculated the root mean squared error of all calibrated test probabilities larger than 0.1. We made this cut off because many of the poorly calibrated models have many points near zero, giving a false low RMSE score. Models without any calibrated probabilities larger than 0.1 were given an rmse of 'N/A'.

4.4. *Feature Collapsing*

Reduction of feature size is also essential for reproducibility of the models in other hospital settings. Our approach collapses features together by utilizing the hierarchical structure of ICD9, ICD10, and CPT4 diagnosis codes and procedure codes, and ATC ingredient definitions for medications. We mapped ICD 9 and 10 diagnosis and procedure codes and drug prescription order ingredients to SNOMED CT concept IDs. About 40% of the features could not be mapped automatically to SNOMED concepts, so we we kept codes and ingredients that did not have concept ids as separate features. We then mapped ICD9 codes to ICD10 codes using General Equivalence Mappings (cite). We reduced all ICD10 diagnosis and procedure code features to the block, subchapter, and chapter (21) level defined by CMS.gov. We then mapped drug prescription ingredients to ATC codes by matching corresponding RxNorm codes to ATC codes. Some drugs required manual mapping to ATC codes. All in all, we successfully mapped 70% of 45 million prescription orders to ATC codes. We then reduced prescription order features to the 5 different levels of ATC description, which include, in increasing order of specificity, anatomical systems (14), therapeutic subgroups (78), pharmacological subgroups

(159), chemical subgroups (364), and chemical substances (895). We then reran the training sets with collapsed features in three combinations: chapters with chemical substrates (the most granular), level 2 with pharmacological subgroups, and level 1 with anatomical systems (the least granular). We chose these pairings to maintain similar ratios of drugs and ICD9 and ICD10 codes. We evaluated the robustness of hierarchical feature collapsing using the method described above. We then compared performance across levels of granularity using AUC-ROC and average precision score, which is a proxy for area under the precision-recall curve.

4.5. *Generalized Linear Model to determine Feature Category Contribution to Models*

To further explore the contribution of conditions, procedures, drugs, and demographics, to the models' classification performance, we performed a multivariate generalized linear model (GLM) on the stroke probabilities generated from each feature category. We retrained the models solely using features from each category. We also added a fifth category, which comprised the ICD9 codes that make up the T-L criteria and ICD10 equivalent codes (see table 1). We added this category to compare the performance of the T-L criteria to the other features in AIS patient classification because our models do not include the T-L criteria in training. We then generated AIS probabilities for the hold out test set using each of the retrained models. Finally, we combined the model probabilities from each of the five models and ran the generalized linear model with binomial distribution to determine which categories significantly (P¡.05) contributed to case determination and their corresponding beta coefficients

4.6. *Additional Patients predicted from EHR*

We chose a probability threshold at the maximum F1 score for each model and identified the number of patients recovered as acute ischemic Stroke patients. We also determined the percentage of these AIS classified patients with the AIS codes defined by T-L by querying

CUIMC's CDM.

References

1. E. J. Benjamin, S. S. Virani, C. W. Callaway, A. M. Chamberlain, A. R. Chang, S. Cheng, S. E. Chiuve, M. Cushman, F. N. Delling, R. Deo and et al., *Circulation* **137**, p. e67e492 (Mar 2018).
2. A. Boehme, C. Esenwa and M. Elkind, *Circulation research* **120**, p. 472495 (2017).
3. H. Markus and S. Bevan, *Nature Reviews Neurology* **10**, p. 723730 (2014).
4. G. Jickling, B. Stamova, B. Ander, X. Zhan, D. Liu, S.-M. Sison, P. Verro and F. Sharp, *Stroke* **43**, p. 20362041 (2012).
5. A. Nouh, M. Hussain, T. Mehta and S. Yaghi, *Frontiers in Neurology* **7**, p. 37 (2016).
6. A. E. Arch, D. C. Weisman, S. Coca, K. V. Nystrom, C. R. Wira and J. L. Schindler, *Stroke* **47**, p. 66873 (Mar 2016).
7. H. Mo, W. Thompson, L. Rasmussen, J. Pacheco, G. Jiang, R. Kiefer, Q. Zhu, J. Xu, E. Montague, D. Carrell and et al., *Journal of the American Medical Informatics Association* **22**, p. 12201230 (2015).
8. C. Shivade, P. Raghavan, E. Fosler-Lussier, P. Embi, N. Elhadad, S. Johnson and A. Lai, *Journal of the American Medical Informatics Association* **21**, p. 221230 (2014).
9. G. Hripcsak and D. J. Albers, *Journal of the American Medical Informatics Association: JAMIA* **20**, p. 11721 (Jan 2013).
10. R. Carroll, A. Eyler and J. Denny, *AMIA ... Annual Symposium proceedings / AMIA Symposium. AMIA Symposium* **2011**, p. 18996 (2011).
11. Y. Chen, R. Carroll, E. Hinz, A. Shah, A. Eyler, J. Denny and H. Xu, *Journal of the American Medical Informatics Association* **20**, p. e253e259 (2013).
12. S. Yu, A. Chakrabortty, K. P. Liao, T. Cai, A. N. Ananthakrishnan, V. S. Gainer, S. E. Churchill, P. Szolovits, S. N. Murphy, I. S. Kohane and et al., *Journal of the American Medical Informatics Association: JAMIA* **24**, p. e143e149 (Apr 2017).
13. W. Ning, S. Chan, A. Beam, M. Yu, A. Geva, K. Liao, M. Mullen, K. D. Mandl, I. Kohane, T. Cai and et al., *Journal of biomedical informatics* , p. 103122 (Feb 2019).
14. P. Peissig, V. Costa, M. Caldwell, C. Rottscheit, R. Berg, E. Mendonca and D. Page, *Journal of*

Biomedical Informatics **52**, p. 260270 (2014).

15. Y. Ni, K. Alwell, C. J. Moomaw, D. Woo, O. Adeoye, M. L. Flaherty, S. Ferioli, J. Mackey, F. De Los Rios La Rosa, S. Martini and et al., *PloS one* **13**, p. e0192586 (Jan 2018).
16. S. Yu, Y. Ma, J. Gronsbell, T. Cai, A. Ananthakrishnan, V. Gainer, S. Churchill, P. Szolovits, S. Murphy, I. Kohane and et al., *Journal of the American Medical Informatics Association: JAMIA* **25** (2017).
17. V. Agarwal, T. Podchiyska, J. M. Banda, V. Goel, T. I. Leung, E. P. Minty, T. E. Sweeney, E. Gyang and N. H. Shah, *Journal of the American Medical Informatics Association: JAMIA* **23**, p. 11661173 (Nov 2016).
18. Y. Halpern, S. Horng, Y. Choi and D. Sontag, *Journal of the American Medical Informatics Association: JAMIA* **23**, p. 73140 (Jul 2016).
19. A. Perotte, R. Pivovarov, K. Natarajan, N. Weiskopf, F. Wood and N. Elhadad, *Journal of the American Medical Informatics Association* **21**, p. 231237 (2014).
20. Y. Zhang, *Association for Computational Linguistics* , p. 6772 (2008).
21. T. F. Imran, D. Posner, J. Honerlaw, J. L. Vassy, R. J. Song, Y.-L. L. Ho, S. J. Kittner, K. P. Liao, T. Cai, C. J. ODonnell and et al., *Clinical epidemiology* **10**, p. 15091521 (Jan 2018).
22. V. Abedi, N. Goyal, G. Tsivgoulis, N. Hosseinichimeh, R. Hontecillas, J. Bassaganya-Riera, L. Elijovich, J. E. Metter, A. W. Alexandrov, D. S. Liebeskind and et al., *Stroke* **48**, p. 16781681 (Jun 2017).
23. Z. Chen, R. Zhang, F. Xu, X. Gong, F. Shi, M. Zhang and M. Lou, *Frontiers in aging neuroscience* **10**, p. 181 (Jan 2018).
24. D. Tirschwell and W. Longstreth, *Stroke* **33**, p. 24652470 (2002).
25. J. A. Sinnott, W. Dai, K. P. Liao, S. Y. Shaw, A. N. Ananthakrishnan, V. S. Gainer, E. W. Karlson, S. Churchill, P. Szolovits, S. Murphy and et al., *Human genetics* **133**, p. 136982 (Nov 2014).
26. J. Sinnott, F. Cai, S. Yu, B. Hejblum, C. Hong, I. Kohane and K. Liao, *Journal of the American Medical Informatics Association: JAMIA* (2018).
27. L. Bastarache, J. Hughey, S. Hebbring, J. Marlo, W. Zhao, W. Ho, S. Driest, T. McGregor, J. Mosley, Q. Wells and et al., *Science* **359**, p. 12331239 (2018).

28. J. H. Son, G. Xie, C. Yuan, L. Ena, Z. Li, A. Goldstein, L. Huang, L. Wang, F. Shen, H. Liu and et al., *American journal of human genetics* **103**, p. 5873 (Jul 2018).
29. G. Hripcsak and D. Albers, *Journal of the American Medical Informatics Association* (2017).
30. G. Hripcsak, M. E. Levine, N. Shang and P. B. Ryan, *Journal of the American Medical Informatics Association: JAMIA* (Nov 2018).
31. F. C. G. C. Polubriaginof, R. Vanguri, K. Quinnies, G. M. Belbin, A. Yahi, H. Salmasian, T. Lorberbaum, V. Nwankwo, L. Li, M. M. Shervey and et al., *Cell* **173**, p. 16921704.e11 (Jun 2018).
32. C. for Medicaid and M. Services, 2018 icd-10 cm and gems `https://www.cms.gov/medicare/coding/icd10/2018-icd-10-cm-and-gems.html`.

Case/ Control Combo	LR Sens	RF Sens	AB Sens	GB Sens	EN Sens	LR Spec	RF Spec	AB Spec	GB Spec	EN Spec	LR PPV	RF PPV	AB PPV	GB PPV	EN PPV	LR NPV	RF NPV	AB NPV	GB NPV	EN NPV
SN	0.9	0.484	0.852	0.738	0.837	0.995	0.999	0.996	0.999	0.998	0.009	0.005	0.008	0.007	0.008	0.999	0.995	0.999	0.997	0.998
SI	0.717	0.408	0.662	0.506	0.713	0.999	1	1	1	0.999	0.007	0.004	0.007	0.005	0.007	0.997	0.994	0.997	0.995	0.997
SC	0.853	0.558	0.89	0.837	0.892	0.997	0.999	0.997	0.998	0.995	0.008	0.006	0.009	0.008	0.009	0.999	0.996	0.999	0.998	0.999
SCI	0.488	0.483	0.464	0.454	0.592	1	1	1	0.999	1	0.005	0.005	0.005	0.005	0.006	0.995	0.995	0.995	0.995	0.996
SR	0.546	0.604	0.508	0.578	0.656	1	1	1	1	1	0.005	0.006	0.005	0.006	0.006	0.995	0.996	0.995	0.996	0.997
TN	0.884	0.472	0.892	0.696	0.752	0.899	0.947	0.946	0.908	0.987	0.01	0.005	0.009	0.008	0.008	0.999	0.994	0.999	0.997	0.998
TI	0.712	0.683	0.701	0.65	0.662	0.999	0.999	0.999	0.999	1	0.007	0.007	0.007	0.006	0.007	0.997	0.997	0.997	0.997	0.997
TC	0.768	0.798	0.789	0.829	0.662	0.999	0.998	0.999	0.996	1	0.008	0.008	0.008	0.008	0.007	0.998	0.998	0.998	0.998	0.997
TCI	0.182	0.309	0.335	0.36	0.172	1	1	0.993	0.999	1	0.002	0.003	0.003	0.004	0.002	0.992	0.993	0.993	0.994	0.992
TR	0.55	0.297	0.524	0.527	0.522	0.999	0.998	0.999	0.998	1	0.005	0.003	0.005	0.005	0.005	0.996	0.993	0.995	0.995	0.995
CN	0.933	0.295	0.955	0.697	0.709	0.86	0.922	0.866	0.887	0.925	0.011	0.003	0.011	0.008	0.008	0.999	0.992	0.999	0.997	0.997
CI	0.204	0.414	0.398	0.58	0.266	1	1	0.993	0.997	1	0.002	0.004	0.004	0.006	0.003	0.992	0.994	0.994	0.996	0.993
CC	0.831	0.818	0.847	0.793	0.769	0.999	0.996	0.998	0.997	0.999	0.008	0.008	0.008	0.008	0.008	0.998	0.998	0.998	0.998	0.998
CCI	0.242	0.373	0.239	0.659	0.338	1	1	0.999	0.99	1	0.002	0.004	0.002	0.007	0.003	0.992	0.994	0.992	0.997	0.993
CR	0.661	0.283	0.52	0.344	0.633	0.996	0.989	0.994	0.996	0.997	0.007	0.003	0.005	0.003	0.006	0.997	0.993	0.995	0.993	0.996

Table 3: Sensitivity, Specificity, Positive Predictive Value, and Negative Predictive Value of Models on Hold-out Test set. Blue values > 0.85, yellow values between 0.65 and 0.85, red values < 0.65.

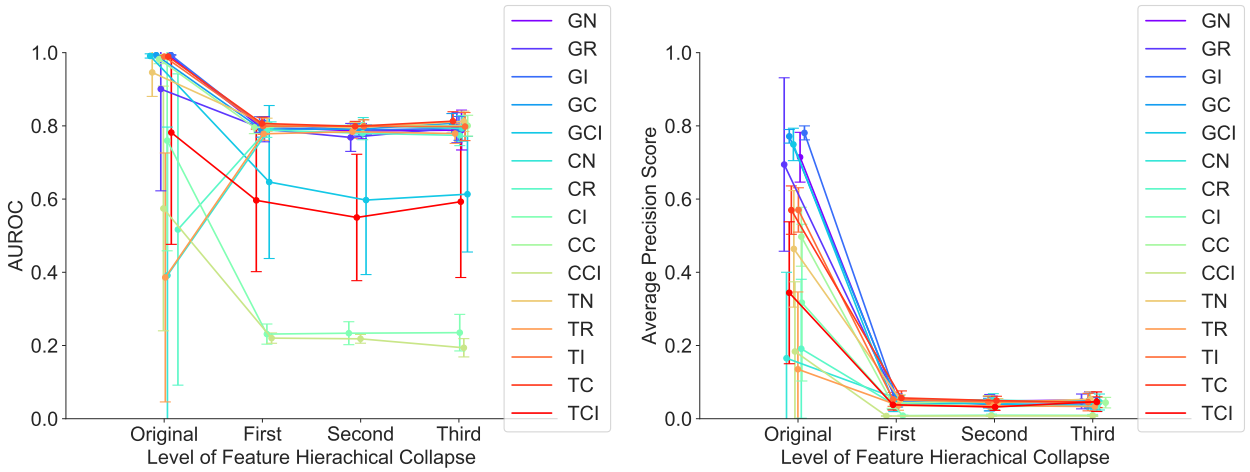
Feature	% Models	Ave. Freq. Cases	Ave. Freq. Controls
CT head or brain; no contrast (2211327)	60%	55%	17%
Age > 50	59%	86%	53%
Duplex scan of extracranial arteries; bilateral (2313975)	52%	28%	9%
MRI, Brain and Brain stem; no contrast (2211351)	43%	37%	8%
Aspirin (1112807)	37%	62%	20%
Transcranial Doppler study intracranial arteries (2313977)	35%	13%	1%
Atorvostatin (1545958)	29%	49%	13%
MR angiography, head; no contrast (2211345)	31%	31%	4%
Encounter for general adult medical examination without abnormal findings (45552054)	29%	70%	60%
Echocardiogram, transthoracic (2313869)	27%	50%	16%
History of TIA or stroke w/o residual effect (44827313)	29%	20%	2%
Male gender	23%	48%	43%
Intercerebral Hemorrhage (44835946)	23%	1%	1%
Female gender	21%	52%	57%
MR angiography, neck, no contrast (2211348)	17%	28%	3%
Hemiplegia, affected unspecified side (44829966)	16%	12%	2%
Convulsions (44823442)	13%	3%	2%
Iatrogenic cerebrovascular infarction or hemorrhage (44834124)	11%	4%	0.4%

Table 4: Common top 10 features in the models. After each of the 75 models were trained, we counted the number of times each feature was represented as one of the top ten by absolute coefficient weight, for methods like logistic regression, or by feature importance, for methods like random forest. Above are features from this analysis along with the proportion of models in which they were in the top ten (% Models), the average frequency in the cases (Ave. Freq. Cases) and the average frequency in the controls (Ave. Freq. Controls).

Case/ Control Combo	LR EHR Prev	RF EHR Prev	AB EHR Prev	GB EHR Prev	EN EHR Prev	LR with Tirsch	RF with Tirsch	AB with Tirsch	GB with Tirsch	EN with Tirsch
SN	1.5	0.7	2.6	2.3	9.7	27.8	30.4	22.4	19.8	5.7
SI	0.8	1.6	1.3	1.7	2.0	48.8	26.2	39.4	30.0	26.9
SC	1.0	2.0	1.7	1.7	2.0	42.3	22.5	36.0	30.7	27.9
SCI	0.2	0.1	0.2	0.2	0.3	85.9	86.2	83.9	78.4	63.6
SR	0.2	0.2	0.3	0.4	0.6	76.7	62.6	73.2	59.9	50.6
TN	97.9	2.1	99.9	29.8	97.1	< 5.0	14.2	< 5.0	< 5.0	< 5.0
TI	1.4	2.4	1.4	1.9	3.7	45.4	29.7	44.8	39.2	26.6
TC	1.6	2.4	1.9	1.5	2.6	42.0	30.3	39.6	44.4	33.2
TCI	0.1	0.0	0.1	0.1	0.8	94.9	87.0	81.8	90.1	42.2
TR	0.7	0.7	0.8	0.5	1.7	50.9	46.9	49.3	67.4	36.6
CN	99.8	3.6	99.9	34.3	98.7	< 5.0	9.6	< 5.0	< 5.0	<5.0
CI	0.1	0.0	0.1	0.1	0.2	96.8	94.7	82.1	89.4	78.2
CC	2.4	3.4	2.3	5.6	18.2	33.7	24.1	34.8	19.9	8.0
CCI	0.0	0.0	0.0	0.2	1.7	98.7	0.0	80.7	65.7	22.4
CR	1.3	1.0	1.3	0.7	2.0	36.5	34.3	33.7	45.1	32.1

Table 5: Prevalence of acute ischemic stroke patients identified by each classifier across the EHR and proportion of those patients with T-L criteria

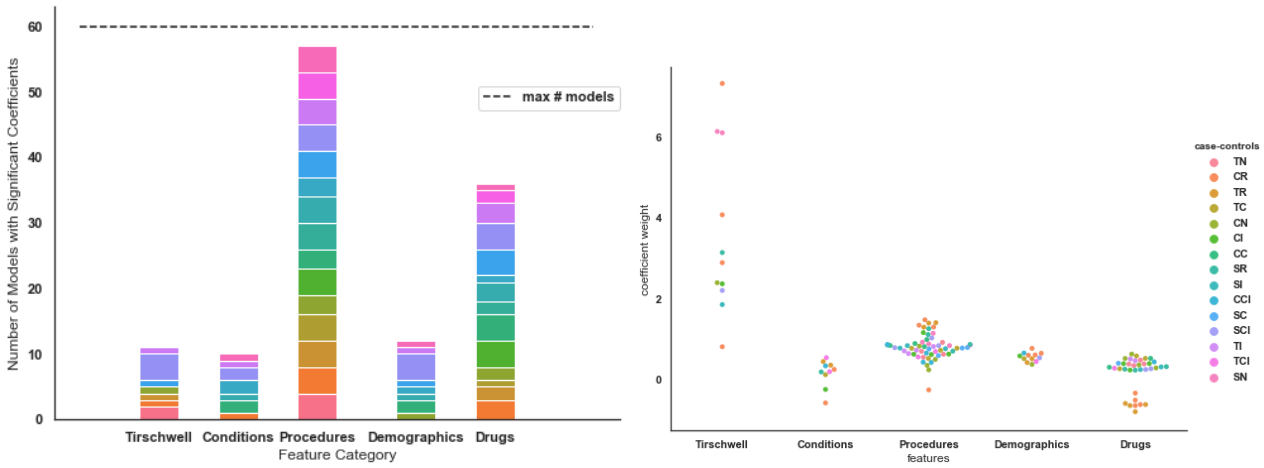
5. Supplementary Materials



(a) AUROC stratified by case-control type

(b) Average Precision Score stratified by case-control type

Fig. 7: Performance of models at increasing level of feature hierarchical collapse. (A) shows AUROC stratified by case-control type, and (B) shows average precision score (AUPR) stratified by case-control type.



(a) Significant feature categories across case-control types

(b) Coefficient weights for feature categories

Fig. 8: Generalized Linear Models differentiate significance across feature types. (A) shows the total number of models with a significant coefficient weight for each feature category, stratified by case-control type. (B) shows the value of the significant coefficient weights, stratified by case-control type

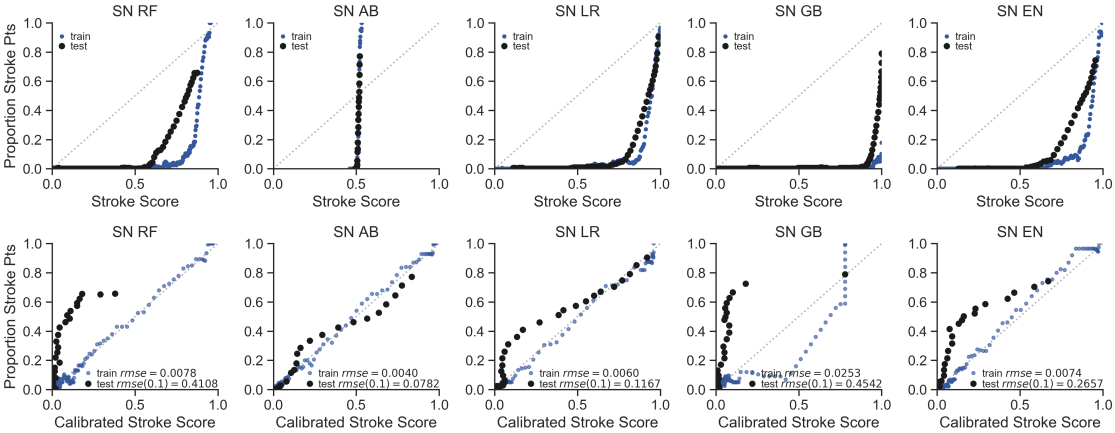


Fig. 9: Classifier type with SN case-control combination varies in calibration success between stroke score and actual proportion of patients at each probability.

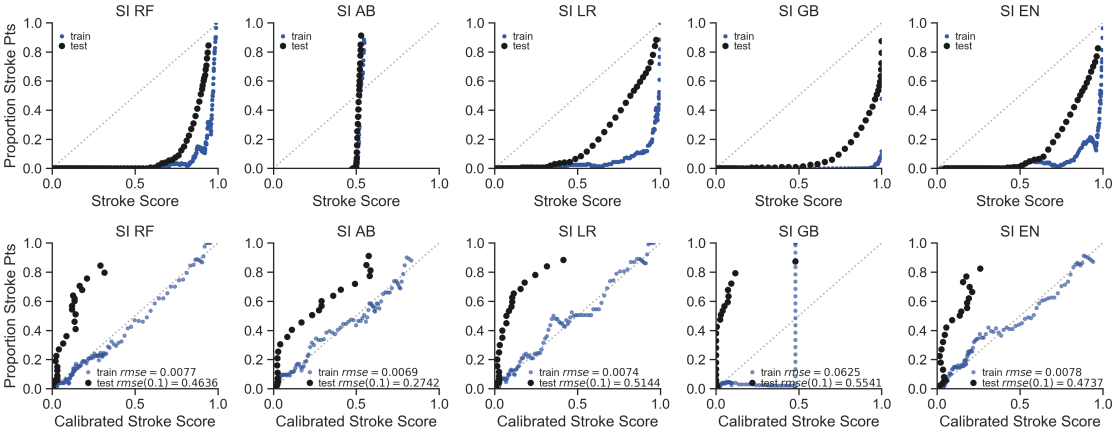


Fig. 10: Classifier type with SI case-control combination varies in calibration success between stroke score and actual proportion of patients at each probability.

		Logistic Regression (L1 Penalty)				Random Forest				Adaboost				Gradient Boosting				Logistic Regression (Elastic Net Penalty)			
	Metric	FB=1	FB=1/2	FB=1/4	FB=1/8	B=1	B=1/2	B=1/4	B=1/8	B=1	B=1/2	B=1/4	B=1/8	B=1	B=1/2	B=1/4	B=1/8	B=1	B=1/2	B=1/4	B=1/8
GN	P	0.67 (0.08)	0.73 (0.15)	0.73 (0.15)	0.73 (0.15)	0.88 (0.06)	0.89 (0.07)	0.89 (0.07)	0.89 (0.07)	0.88 (0.06)	0.89 (0.07)	0.89 (0.07)	0.89 (0.07)	0.75 (0.15)	0.77 (0.16)	0.77 (0.16)	0.77 (0.16)	0.86 (0.09)	0.86 (0.10)	0.86 (0.10)	0.86 (0.10)
	R	0.90 (0.03)	0.80 (0.22)	0.80 (0.22)	0.80 (0.22)	0.48 (0.10)	0.46 (0.12)	0.46 (0.12)	0.46 (0.12)	0.48 (0.10)	0.46 (0.12)	0.46 (0.12)	0.46 (0.12)	0.85 (0.07)	0.80 (0.16)	0.80 (0.16)	0.80 (0.16)	0.74 (0.13)	0.71 (0.18)	0.71 (0.18)	0.71 (0.18)
	F	0.77 (0.04)	0.72 (0.14)	0.72 (0.14)	0.72 (0.14)	0.61 (0)	0.59 (0.10)	0.59 (0.10)	0.59 (0.10)	0.61 (0)	0.59 (0.10)	0.59 (0.10)	0.59 (0.10)	0.78 (0.08)	0.75 (0.11)	0.75 (0.11)	0.75 (0.11)	0.78 (0.06)	0.75 (0.11)	0.75 (0.11)	0.75 (0.11)
	T	0.89 (0.04)	0.91 (0.06)	0.91 (0.06)	0.91 (0.06)	0.87 (0)	0.87 (0.03)	0.87 (0.03)	0.87 (0.03)	0.87 (0)	0.87 (0.03)	0.87 (0.03)	0.87 (0.03)	0.51 (0.00)	0.51 (0.01)	0.51 (0.01)	0.51 (0.01)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
	# >T	1354 (206)	1184 (450)	1184 (450)	1184 (450)	559 (147)	534 (173)	534 (173)	534 (173)	1218 (469)	1152 (533)	1152 (533)	1152 (533)	885 (245)	855 (290)	855 (290)	855 (290)	992 (175)	941 (155)	889 (255)	889 (255)
	# > T in EHR	78354				38189				135886				124452				515807			
GI	P	0.95 (0.06)	0.98 (0.02)	0.99 (0.02)	0.99 (0.02)	0.96 (0.07)	1.00 (0.01)	1.00 (0.01)	1.00 (0.01)	0.96 (0.07)	1.00 (0.01)	1.00 (0.01)	1.00 (0.01)	0.95 (0.05)	0.98 (0.02)	0.98 (0.02)	0.98 (0.02)	0.99 (0.01)	0.99 (0.01)	1.00 (0.00)	1.00 (0.00)
	R	0.72 (0.16)	0.62 (0.14)	0.57 (0.16)	0.57 (0.16)	0.41 (0.21)	0.18 (0.11)	0.16 (0.11)	0.16 (0.11)	0.41 (0.21)	0.18 (0.11)	0.16 (0.11)	0.16 (0.11)	0.66 (0.17)	0.49 (0.24)	0.49 (0.24)	0.49 (0.24)	0.51 (0.17)	0.45 (0.18)	0.24 (0.20)	0.24 (0.20)
	F	0.80 (0.09)	0.75 (0.10)	0.71 (0.13)	0.71 (0.13)	0.53 (0)	0.29 (0.15)	0.26 (0.16)	0.26 (0.16)	0.53 (0)	0.29 (0.15)	0.26 (0.16)	0.26 (0.16)	0.76 (0.11)	0.61 (0.23)	0.61 (0.23)	0.61 (0.23)	0.65 (0.15)	0.59 (0.18)	0.34 (0.24)	0.34 (0.24)
	T	0.95 (0.06)	0.98 (0.02)	0.98 (0.02)	0.98 (0.02)	0.94 (0)	0.97 (0.01)	0.97 (0.01)	0.97 (0.01)	0.94 (0)	0.97 (0.01)	0.97 (0.01)	0.97 (0.01)	0.52 (0.01)	0.53 (0.01)	0.53 (0.01)	0.53 (0.01)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
	# >T	767 (217)	632 (155)	580 (175)	580 (175)	442 (274)	177 (111)	158 (116)	158 (116)	704 (213)	501 (251)	501 (251)	501 (251)	509 (172)	447 (184)	236 (195)	236 (195)	790 (266)	590 (205)	529 (232)	529 (232)
	# > T in EHR	41665				84812				86702				88446				106901			
GC	P	0.80 (0.16)	0.85 (0.12)	0.85 (0.12)	0.85 (0.12)	0.88 (0.14)	0.89 (0.15)	0.89 (0.15)	0.89 (0.15)	0.88 (0.14)	0.89 (0.15)	0.89 (0.15)	0.89 (0.15)	0.79 (0.18)	0.79 (0.18)	0.79 (0.18)	0.79 (0.18)	0.86 (0.15)	0.88 (0.15)	0.88 (0.16)	0.88 (0.16)
	R	0.85 (0.13)	0.82 (0.13)	0.82 (0.13)	0.82 (0.13)	0.56 (0.22)	0.48 (0.27)	0.48 (0.27)	0.48 (0.27)	0.56 (0.22)	0.48 (0.27)	0.48 (0.27)	0.48 (0.27)	0.89 (0.10)	0.89 (0.10)	0.89 (0.10)	0.89 (0.10)	0.84 (0.12)	0.80 (0.14)	0.75 (0.26)	0.75 (0.26)
	F	0.80 (0.07)	0.82 (0.04)	0.82 (0.04)	0.82 (0.04)	0.63 (0)	0.56 (0.19)	0.56 (0.19)	0.56 (0.19)	0.63 (0)	0.56 (0.19)	0.56 (0.19)	0.56 (0.19)	0.81 (0.09)	0.81 (0.09)	0.81 (0.09)	0.81 (0.09)	0.83 (0.07)	0.82 (0.07)	0.75 (0.22)	0.75 (0.22)
	T	0.83 (0.18)	0.90 (0.10)	0.90 (0.10)	0.90 (0.10)	0.94 (0)	0.95 (0.04)	0.95 (0.04)	0.95 (0.04)	0.94 (0)	0.95 (0.04)	0.95 (0.04)	0.95 (0.04)	0.52 (0.01)	0.52 (0.01)	0.52 (0.01)	0.52 (0.01)	0.99 (0.02)	0.99 (0.02)	0.99 (0.02)	0.99 (0.02)
	# >T	1156 (442)	1005 (294)	1005 (294)	1005 (294)	703 (459)	621 (504)	621 (504)	621 (504)	1226 (472)	1226 (472)	1226 (472)	1226 (472)	1031 (369)	978 (389)	925 (466)	925 (466)	1338 (534)	1044 (565)	1044 (565)	1044 (565)
	# > T in EHR	55684				109126				89961				91732				108250			
GCI	P	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	0.88 (0.22)	0.92 (0.20)	0.92 (0.20)	0.92 (0.20)
	R	0.49 (0.11)	0.43 (0.10)	0.43 (0.10)	0.43 (0.10)	0.48 (0.13)	0.33 (0.11)	0.33 (0.11)	0.33 (0.11)	0.48 (0.13)	0.33 (0.11)	0.33 (0.11)	0.33 (0.11)	0.46 (0.10)	0.37 (0.14)	0.31 (0.18)	0.31 (0.18)	0.45 (0.08)	0.39 (0.12)	0.35 (0.15)	0.35 (0.15)
	F	0.65 (0.11)	0.60 (0.10)	0.60 (0.10)	0.60 (0.10)	0.64 (0)	0.49 (0.12)	0.49 (0.12)	0.49 (0.12)	0.64 (0)	0.49 (0.12)	0.49 (0.12)	0.49 (0.12)	0.63 (0.10)	0.52 (0.15)	0.44 (0.21)	0.44 (0.21)	0.58 (0.09)	0.53 (0.15)	0.48 (0.18)	0.48 (0.18)
	T	0.95 (0.03)	0.96 (0.02)	0.96 (0.02)	0.96 (0.02)	0.82 (0)	0.86 (0.03)	0.86 (0.03)	0.86 (0.03)	0.82 (0)	0.86 (0.03)	0.86 (0.03)	0.86 (0.03)	0.51 (0.00)	0.51 (0.00)	0.52 (0.01)	0.52 (0.01)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
	# >T	487 (111)	432 (103)	432 (103)	432 (103)	481 (125)	333 (112)	333 (112)	333 (112)	463 (104)	365 (139)	310 (181)	310 (181)	587 (295)	469 (264)	431 (286)	431 (286)	592 (87)	461 (157)	412 (127)	412 (127)
	# > T in EHR	8684				5089				9918				10915				17267			
GR	P	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	0.97 (0.02)	0.99 (0.01)	1.00 (0.01)	1.00 (0.01)	0.97 (0.02)	0.99 (0.01)	1.00 (0.01)	1.00 (0.01)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	0.99 (0.01)	1.00 (0.01)	1.00 (0.01)	1.00 (0.01)
	R	0.55 (0.15)	0.35 (0.17)	0.35 (0.17)	0.35 (0.17)	0.60 (0.06)	0.38 (0.16)	0.33 (0.16)	0.33 (0.16)	0.60 (0.06)	0.38 (0.16)	0.33 (0.16)	0.33 (0.16)	0.51 (0.13)	0.35 (0.15)	0.35 (0.15)	0.35 (0.15)	0.58 (0.11)	0.35 (0.14)	0.35 (0.14)	0.35 (0.14)
	F	0.69 (0.14)	0.49 (0.18)	0.49 (0.18)	0.49 (0.18)	0.74 (0)	0.53 (0.16)	0.47 (0.18)	0.47 (0.18)	0.74 (0)	0.53 (0.16)	0.47 (0.18)	0.47 (0.18)	0.66 (0.13)	0.50 (0.16)	0.50 (0.16)	0.50 (0.16)	0.72 (0.09)	0.50 (0.15)	0.50 (0.15)	0.50 (0.15)
	T	0.95 (0.03)	0.98 (0.02)	0.98 (0.02)	0.98 (0.02)	0.82 (0)	0.89 (0.04)	0.90 (0.04)	0.90 (0.04)	0.82 (0)	0.89 (0.04)	0.90 (0.04)	0.90 (0.04)	0.51 (0.02)	0.52 (0.00)	0.52 (0.00)	0.52 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
	# >T	546 (150)	348 (168)	348 (168)	348 (168)	620 (68)	377 (161)	328 (163)	328 (163)	508 (135)	348 (145)	348 (145)	348 (145)	579 (112)	348 (140)	348 (140)	348 (140)	661 (142)	534 (144)	444 (162)	444 (162)
	# > T in EHR	12612				11057				15494				22312				33741			
TN	P	0.16 (0.10)	0.39 (0.32)	0.83 (0.21)	0.86 (0.21)	0.10 (0.03)	0.13 (0.01)	0.13 (0.02)	0.13 (0.02)	0.10 (0.03)	0.13 (0.01)	0.13 (0.02)	0.13 (0.02)	0.25 (0.12)	0.55 (0.24)	0.74 (0.20)	0.79 (0.10)	0.12 (0.10)	0.37 (0.18)	0.53 (0.13)	0.62 (0.17)
	R	0.88 (0.03)	0.78 (0.15)	0.50 (0.20)	0.44 (0.23)	0.47 (0.08)	0.28 (0.12)	0.18 (0.13)	0.18 (0.13)	0.47 (0.08)	0.28 (0.12)	0.18 (0.13)	0.18 (0.13)	0.89 (0.03)	0.79 (0.09)	0.66 (0.16)	0.49 (0.25)	0.70 (0.09)	0.49 (0.18)	0.33 (0.15)	0.25 (0.17)
	F	0.26 (0.15)	0.40 (0.26)	0.56 (0.13)	0.50 (0.18)	0.16 (0)	0.17 (0.03)	0.13 (0.06)	0.13 (0.06)	0.16 (0)	0.17 (0.03)	0.13 (0.06)	0.13 (0.06)	0.37 (0.14)	0.60 (0.17)	0.65 (0.13)	0.55 (0.20)	0.19 (0.12)	0.35 (0.11)	0.37 (0.08)	0.29 (0.14)
	T	0.94 (0.01)	0.95 (0.02)	0.98 (0.01)	0.98 (0.01)	0.75 (0)	0.79 (0.03)	0.82 (0.04)	0.82 (0.04)	0.75 (0)	0.79 (0.03)	0.82 (0.04)	0.82 (0.04)	0.51 (0.00)	0.51 (0.00)	0.51 (0.00)	0.51 (0.00)	0.96 (0.02)	0.98 (0.01)	0.99 (0.01)	0.99 (0.01)
	# >T	10984 (9517)	9517 (10639)	734 (552)	647 (584)	5760 (3250)	2267 (1153)	1323 (1023)	1323 (1023)	6245 (7751)	2037 (1410)	1167 (1088)	645 (368)	9861 (6113)	2797 (4143)	736 (522)	513 (501)	2002 (987)	1143 (790)	269 (223)	181 (207)
	# > T in EHR	>1000000				113191				>1000000				>1000000				>1000000			
TI	P	0.94 (0.06)	0.98 (0.02)	1.00 (0.01)	1.00 (0.01)	0.90 (0.07)	0.94 (0.05)	0.99 (0.01)	0.99 (0.01)	0.90 (0.07)	0.94 (0.05)	0.99 (0.01)	0.99 (0.01)	0.89 (0.07)	0.97 (0.03)	0.99 (0.01)	0.99 (0.01)	0.84 (0.08)	0.90 (0.03)	0.92 (0.03)	0.96 (0.04)
	R	0.71 (0.13)	0.56 (0.18)	0.20 (0.17)	0.17 (0.17)	0.68 (0.13)	0.56 (0.16)	0.30 (0.12)	0.24 (0.15)	0.68 (0.13)	0.56 (0.16)	0.30 (0.12)	0.24 (0.15)	0.70 (0.09)	0.39 (0.19)	0.18 (0.11)	0.16 (0.11)	0.65 (0.16)	0.47 (0.13)	0.37 (0.12)	0.19 (0.17)
	F	0.80 (0.08)	0.69 (0.15)	0.31 (0.20)	0.26 (0.20)	0.76 (0)	0.68 (0.10)	0.45 (0.15)	0.36 (0.20)	0.76 (0)	0.68 (0.10)	0.45 (0.15)	0.36 (0.20)	0.78 (0.03)	0.52 (0.20)	0.30 (0.15)	0.25 (0.16)	0.71 (0.09)	0.60 (0.09)	0.51 (0.13)	0.28 (0.22)
	T	0.96 (0.04)	0.98 (0.02)	1.00 (0.00)	1.00 (0.00)	0.91 (0)	0.93 (0.02)	0.95 (0.01)	0.96 (0.01)	0.91 (0)	0.93 (0.02)	0.95 (0.01)	0.96 (0.01)	0.51 (0.00)	0.52 (0.01)	0.52 (0.01)	0.53 (0.01)	0.99 (0.01)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
	# >T	765 (187)	573 (199)	205 (170)	167 (169)	770 (207)	603 (212)	304 (121)	242 (154)	800 (171)	404 (212)	184 (109)	156 (113)	797 (275)	524 (162)	404 (135)	202 (184)	689 (177)	470 (263)	95 (94)	52 (56)
	# > T in EHR	73580				127117				75240				103072				196063			
TC	P	0.93 (0.04)	0.98 (0.02)	0.99 (0.01)	0.99 (0.01)	0.85 (0.08)	0.92 (0.09)	0.97 (0.05)	0.97 (0.05)	0.85 (0.08)	0.92 (0.09)	0.97 (0.05)	0.97 (0.05)	0.91 (0.04)	0.96 (0.03)	0.96 (0.03)	0.96 (0.03)	0.69 (0.10)	0.82 (0.09)	0.85 (0.06)	0.85 (0.06)
	R	0.77 (0.09)	0.61 (0.12)	0.55 (0.14)	0.55 (0.14)	0.80 (0.11)	0.64 (0.16)	0.38 (0.22)	0.38 (0.22)	0.80 (0.11)	0.64 (0.16)	0.38 (0.22)	0.38 (0.22)	0.79 (0.08)	0.66 (0.10)	0.64 (0.12)	0.64 (0.12)	0.83 (0.08)	0.64 (0.13)	0.57 (0.14)	0.57 (0.14)
	F	0.84 (0.04)	0.75 (0.09)	0.69 (0.11)	0.69 (0.11)	0.81 (0)	0.73 (0.09)	0.50 (0.21)	0.50 (0.21)	0.81 (0)	0.73 (0.09)	0.50 (0.21)	0.50 (0.21)	0.84 (0.04)	0.78 (0.07)	0.76 (0.08)	0.76 (0.08)	0.74 (0.04)	0.71 (0.05)	0.67 (0.09)	0.67 (0.09)
	T	0.97 (0.02)	0.99 (0.01)	0.99 (0.01)	0.99 (0.01)	0.90 (0)	0.92 (0.03)	0.95 (0.02)	0.95 (0.02)	0.90 (0)	0.92 (0.03)	0.95 (0.02)	0.95 (0.02)	0.51 (0.00)	0.52 (0.00)	0.52 (0.00)	0.52 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
	# >T	826 (130)	629 (135)	557 (148)	557 (148)	951 (213)	717 (275)	399 (268)	399 (268)	866 (115)	694 (117)	671 (137)	671 (137)	1251 (302)	805 (258)	680 (222)	680 (222)	686 (133)	464 (194)	341 (235)	278 (228)
	# > T in EHR	86952				128788				103380				81746				137181			
TCI	P	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	0.93 (0.04)	0.97 (0.01)	0.97 (0.01)	0.86 (0.29)	0.93 (0.04)	0.97 (0.01)	0.97 (0.01)	0.86 (0.29)	0.37 (0.15)	0.54 (0.19)	0.52 (0.19)	0.54 (0.18)	0.75 (0.07)	0.83 (0.05)	0.88 (0.04)	0.91 (0.06)
	R	0.18 (0.07)	0.07 (0.04)	0.04 (0.03)	0.04 (0.03)	0.31 (0.09)	0.12 (0.04)	0.10 (0.05)	0.09 (0.06)	0.31 (0.09)	0.12 (0.04)	0.10 (0.05)	0.09 (0.06)	0.33 (0.04)	0.20 (0.07)	0.13 (0.06)	0.12 (0.06)	0.36 (0.08)	0.24 (0.06)	0.18 (0.07)	0.12 (0.06)
	F	0.30 (0.11)	0.12 (0.06)	0.08 (0.05)	0.07 (0.05)	0.45 (0)	0.21 (0.07)	0.18 (0.08)	0.15 (0.11)	0.45 (0)	0.21 (0.07)	0.18 (0.08)	0.15 (0.11)	0.33 (0.04)	0.27 (0.05)	0.20 (0.07)	0.19 (0.07)	0.48 (0.06)	0.37 (0.07)	0.29 (0.10)	0.20 (0.10)
	T	0.87 (0.04)	0.93 (0.02)	0.95 (0.02)	0.95 (0.02)	0.72 (0)	0.78 (0.02)	0.79 (0.02)	0.81 (0.04)	0.72 (0)	0.78 (0.02)	0.79 (0.02)	0.81 (0.04)	0.50 (0.00)	0.51 (0.00)	0.51 (0.00)	0.51 (0.00)	0.95 (0.02)	0.98 (0.01)	0.99 (0.01)	0.99 (0.01)
	# >T	181 (75)	65 (37)	42 (28)	36 (29)	335 (105)	123 (46)	106 (52)	89 (66)	1051 (413)	512 (470)	299 (248)	261 (239)	487 (126)	294 (75)	207 (84)	131 (78)	172 (50)	77 (61)	28 (20)	19 (16)
	# > T in EHR	4126				1468				5997				6221				42254			
TR	P	0.88 (0.09)	0.92 (0.11)	1.00 (0.01)	1.00 (0.00)	0.67 (0.10)	0.90 (0.05)	0.93 (0.05)	0.95 (0.03)	0.67 (0.10)	0.90 (0.05)	0.93 (0.05)	0.95 (0.03)	0.84 (0.09)	0.88 (0.09)	0.97 (0.05)	1.00 (0.01)	0.83 (0.15)	0.96 (0.05)	1.00 (0.01)	0.90 (0.30)
	R	0.55 (0.12)	0.41 (0.22)	0.09 (0.09)	0.03 (0.03)	0.30 (0.07)	0.12 (0.05)	0.09 (0.04)	0.07 (0.04)	0.30 (0.07)	0.12 (0.05)	0.09 (0.04)	0.07 (0.04)	0.52 (0.12)	0.39 (0.20)	0.08 (0.06)	0.02 (0.02)	0.53 (0.14)	0.31 (0.16)	0.10 (0.10)	0.03 (0.03)
	F	0.66 (0.08)	0.51 (0.19)	0.15 (0.14)	0.06 (0.06)	0.40 (0)	0.20 (0.07)	0.17 (0.07)	0.13 (0.07)	0.40 (0)	0.20 (0.07)	0.17 (0.07)	0.13 (0.07)	0.63 (0.10)	0.49 (0.18)	0.14 (0.10)	0.03 (0.04)	0.61 (0.06)	0.44 (0.17)	0.16 (0.13)	0.06 (0.05)
	T	0.86 (0.04)	0.90 (0.06)	0.97 (0.02)	0.99 (0.01)	0.82 (0)	0.89 (0.02)	0.90 (0.02)	0.91 (0.03)	0.82 (0)	0.89 (0.02)	0.90 (0.02)	0.91 (0.03)	0.50 (0.00)	0.51 (0.00)	0.51 (0.00)	0.51 (0.00)	0.94 (0.04)	0.98 (0.01)	0.99 (0.01)	1.00 (0.00)
	# >T	641 (206)	476 (309)	90 (91)	31 (34)	469 (165)	133 (57)	105 (57)	75 (50)	639 (193)	462 (283)	82 (68)	16 (19)	705 (411)	330 (193)	97 (101)	33 (29)	564 (176)	324 (206)	85 (93)	40 (80)
	# > T in EHR	39586				35727				42168				24736				91407			
CN	P	0.11 (0.09)	0.47 (0.24)	0.65 (0.25)	0.81 (0.18)	0.04 (0.01)	0.04 (0.01)	0.04 (0.01)	0.04 (0.01)	0.04 (0.01)	0.04 (0.01)	0.04 (0.01)	0.04 (0.01)	0.11 (0.06)	0.23 (0.13)	0.36 (0.17)	0.53 (0.19)	0.11 (0.07)	0.21 (0.06)	0.29 (0.07)	0.31 (0.08)
	R	0.93 (0.02)	0.81 (0.10)	0.71 (0.13)	0.57 (0.20)	0.29 (0.11)	0.17 (0.09)	0.09 (0.08)	0.09 (0.08)	0.29 (0.11)	0.17 (0.09)	0.09 (0.08)	0.09 (0.08)	0.95 (0.01)	0.89 (0.06)	0.78 (0.17)	0.53 (0.28)	0.70 (0.10)	0.51 (0.15)	0.32 (0.14)	0.29 (0.15)
	F	0.19 (0.13)	0.53 (0.19)	0.62 (0.13)	0.61 (0.15)	0.07 (0)	0.06 (0.02)	0.04 (0.02)	0.04 (0.02)	0.07 (0)	0.06 (0.02)	0.04 (0.02)	0.04 (0.02)	0.19 (0.09)	0.34 (0.16)	0.44 (0.11)	0.43 (0.11)	0.18 (0.10)	0.28 (0.05)	0.28 (0.04)	0.26 (0.06)
	T	0.95 (0.01)	0.97 (0.01)	0.97 (0.01)	0.98 (0.01)	0.75 (0)	0.79 (0.04)	0.82 (0.04)	0.82 (0.04)	0.75 (0)	0.79 (0.04)	0.82 (0.04)	0.82 (0.04)	0.51 (0.00)	0.51 (0.00)	0.51 (0.00)	0.51 (0.00)	0.97 (0.01)	0.99 (0.01)	0.99 (0.00)	0.99 (0.00)
	# >T	14885 (9502)	3274 (3773)	1594 (1489)	822 (572)	8124 (4352)	3767 (2162)	2446 (2398)	2446 (2398)	14324 (10536)	7492 (8353)	2892 (1741)	1471 (1426)	11990 (9939)	2924 (1844)	1238 (743)	1126 (808)	8205 (8386)	2612 (4339)	329 (247)	149 (132)
	# > T in EHR	>1000000				190859				>1000000				>1000000				>1000000			
CI	P	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	0.99 (0.00)	0.99 (0.00)	0.99 (0.01)	0.99 (0.01)	0.99 (0.00)	0.99 (0.00)	0.99 (0.01)	0.99 (0.01)	0.50 (0.27)	0.83 (0.16)	0.93 (0.09)	0.93 (0.09)	0.64 (0.05)	0.72 (0.07)	0.78 (0.04)	0.81 (0.06)
	R	0.20 (0.08)	0.12 (0.04)	0.07 (0.04)	0.05 (0.04)	0.41 (0.09)	0.21 (0.06)	0.12 (0.06)	0.09 (0.07)	0.41 (0.09)	0.21 (0.06)	0.12 (0.06)	0.09 (0.07)	0.40 (0.04)	0.32 (0.03)	0.24 (0.05)	0.23 (0.06)	0.58 (0.07)	0.43 (0.13)	0.33 (0.09)	0.25 (0.11)
	F	0.33 (0.11)	0.21 (0.07)	0.12 (0.06)	0.10 (0.06)	0.58 (0)	0.34 (0.08)	0.21 (0.10)	0.16 (0.11)	0.58 (0)	0.34 (0.08)	0.21 (0.10)	0.16 (0.11)	0.40 (0.06)	0.46 (0.03)	0.37 (0.06)	0.36 (0.06)	0.60 (0.02)	0.52 (0.08)	0.46 (0.08)	0.36 (0.12)
	T	0.84 (0.06)	0.89 (0.04)	0.94 (0.02)	0.94 (0.03)	0.60 (0)	0.66 (0.01)	0.68 (0.02)	0.70 (0.03)	0.60 (0)	0.66 (0.01)	0.68 (0.02)	0.70 (0.03)	0.50 (0.00)	0.51 (0.00)	0.51 (0.00)	0.51 (0.00)	0.93 (0.03)	0.97 (0.02)	0.99 (0.01)	0.99 (0.01)
	# >T	204 (82)	121 (43)	66 (36)	54 (36)	419 (94)	211 (59)	124 (65)	93 (69)	1047 (474)	417 (136)	259 (90)	250 (95)	916 (179)	615 (245)	435 (129)	318 (164)	266 (144)	130 (61)	76 (59)	49 (40)
	# > T in EHR	3532				19				5399				4322				11205			
CC	P	0.85 (0.04)	0.90 (0.04)	0.94 (0.04)	0.98 (0.04)	0.67 (0.05)	0.79 (0.11)	0.95 (0.04)	0.96 (0.04)	0.67 (0.05)	0.79 (0.11)	0.95 (0.04)	0.96 (0.04)	0.83 (0.07)	0.91 (0.05)	0.93 (0.06)	0.97 (0.06)	0.72 (0.08)	0.79 (0.06)	0.92 (0.05)	0.94 (0.04)
	R	0.83 (0.03)	0.78 (0.06)	0.64 (0.15)	0.44 (0.21)	0.82 (0.05)	0.63 (0.18)	0.26 (0.16)	0.22 (0.16)	0.82 (0.05)	0.63 (0.18)	0.26 (0.16)	0.22 (0.16)	0.85 (0.04)	0.75 (0.08)	0.61 (0.21)	0.36 (0.23)	0.79 (0.05)	0.69 (0.12)	0.34 (0.19)	0.25 (0.14)
	F	0.84 (0.01)	0.83 (0.02)	0.75 (0.10)	0.57 (0.19)	0.73 (0)	0.67 (0.08)	0.37 (0.19)	0.33 (0.20)	0.73 (0)	0.67 (0.08)	0.37 (0.19)	0.33 (0.20)	0.83 (0.02)	0.82 (0.03)	0.71 (0.15)	0.48 (0.22)	0.75 (0.03)	0.73 (0.07)	0.46 (0.18)	0.37 (0.16)
	T	0.98 (0.01)	0.99 (0.01)	0.99 (0.01)	1.00 (0.00)	0.90 (0)	0.92 (0.02)	0.95 (0.01)	0.96 (0.01)	0.90 (0)	0.92 (0.02)	0.95 (0.01)	0.96 (0.01)	0.51 (0.00)	0.52 (0.00)	0.52 (0.01)	0.53 (0.01)	0.99 (0.01)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
	# >T	974 (77)	869 (109)	688 (186)	452 (234)	1234 (174)	837 (337)	275 (176)	235 (174)	1032 (143)	828 (132)	672 (267)	386 (288)	1120 (203)	888 (207)	376 (232)	273 (173)	888 (239)	723 (236)	492 (271)	233 (252)
	# > T in EHR	126804				181864				121005				298356				964909			
CCI	P	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	0.97 (0.02)	0.98 (0.01)	0.99 (0.01)	0.98 (0.02)	0.97 (0.02)	0.98 (0.01)	0.99 (0.01)	0.98 (0.02)	0.81 (0.14)	0.76 (0.19)	0.83 (0.09)	0.87 (0.09)	0.42 (0.09)	0.61 (0.07)	0.68 (0.10)	0.72 (0.09)
	R	0.24 (0.05)	0.11 (0.03)	0.06 (0.03)	0.05 (0.03)	0.37 (0.06)	0.21 (0.08)	0.15 (0.04)	0.12 (0.07)	0.37 (0.06)	0.21 (0.08)	0.15 (0.04)	0.12 (0.07)	0.24 (0

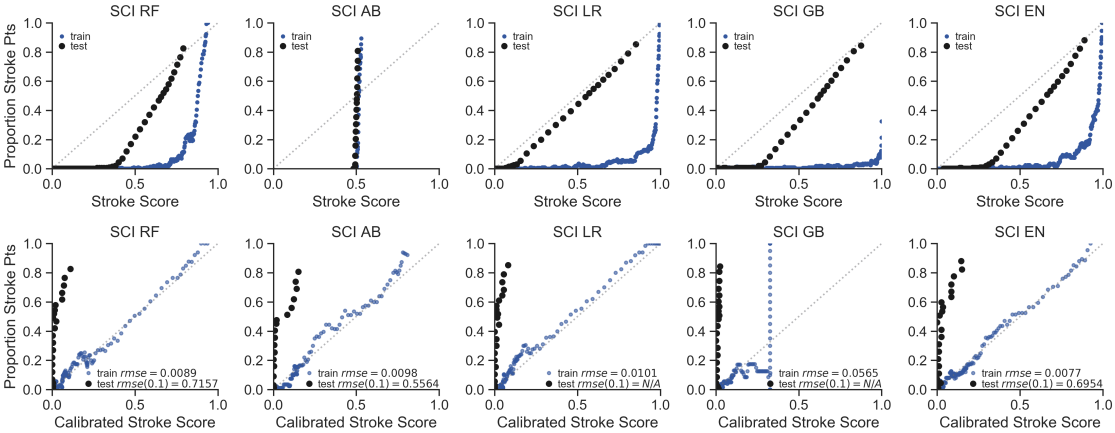


Fig. 11: Classifier type with SCI case-control combination varies in calibration success between stroke score and actual proportion of patients at each probability.

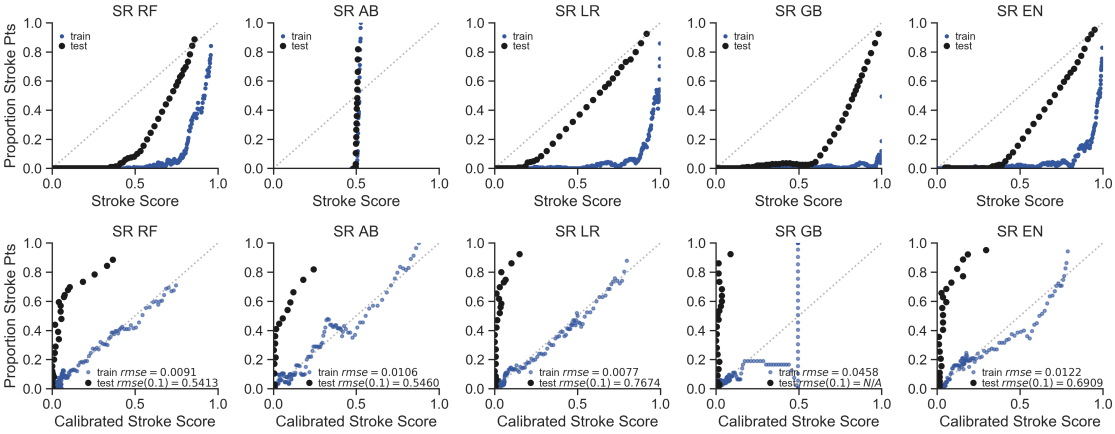


Fig. 12: Classifier type with SR case-control combination varies in calibration success between stroke score and actual proportion of patients at each probability.

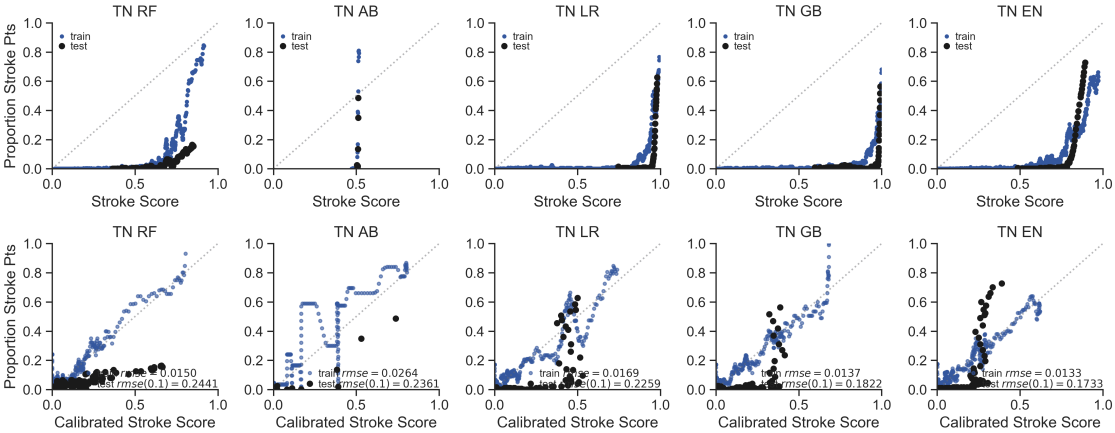


Fig. 13: Classifier type with TN case-control combination varies in calibration success between stroke score and actual proportion of patients at each probability.

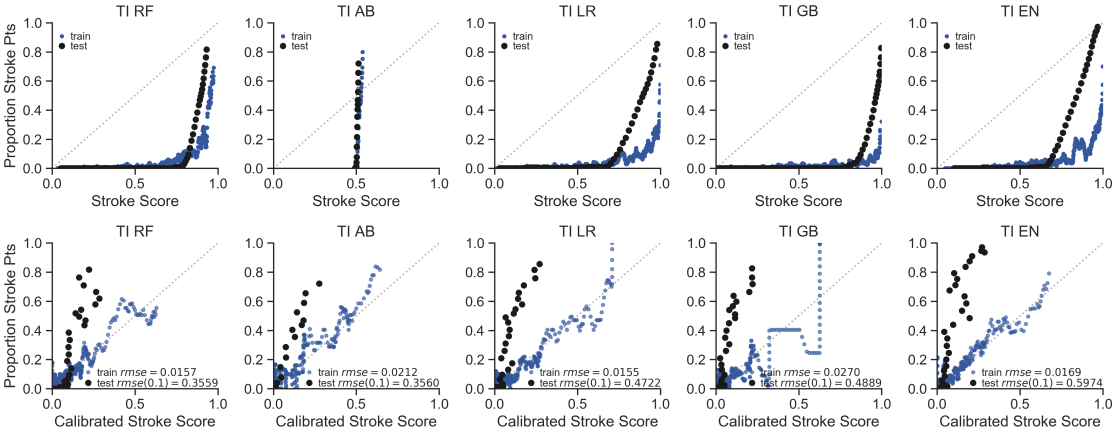


Fig. 14: Classifier type with TI case-control combination varies in calibration success between stroke score and actual proportion of patients at each probability.

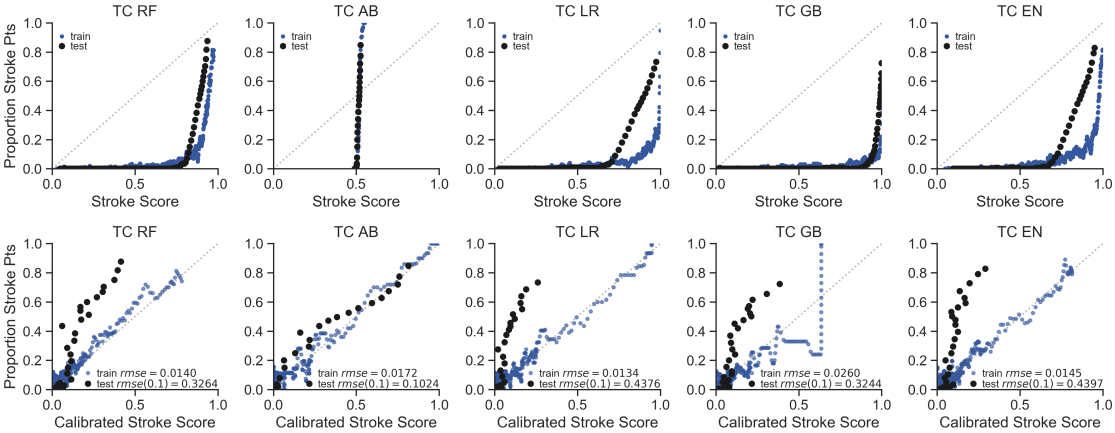


Fig. 15: Classifier type with TC case-control combination varies in calibration success between stroke score and actual proportion of patients at each probability.

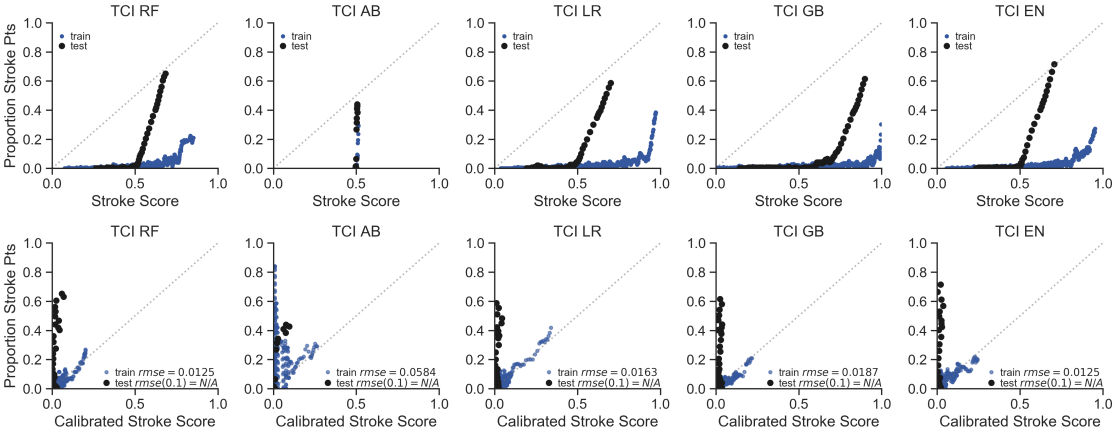


Fig. 16: Classifier type with TCI case-control combination varies in calibration success between stroke score and actual proportion of patients at each probability.

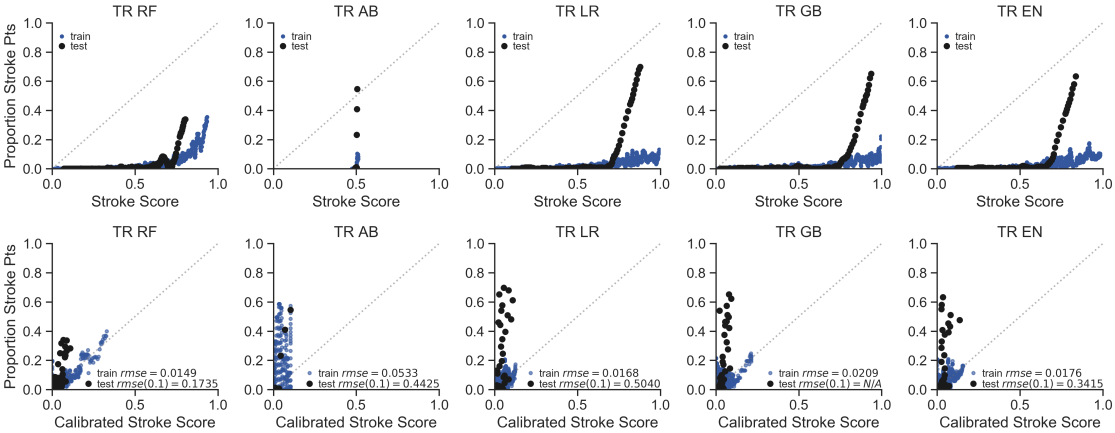


Fig. 17: Classifier type with TR case-control combination varies in calibration success between stroke score and actual proportion of patients at each probability.

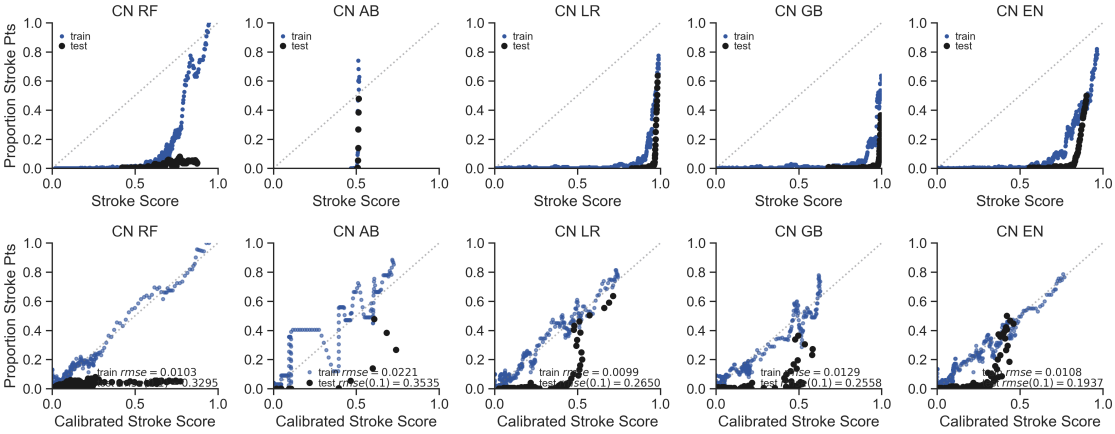


Fig. 18: Classifier type with CN case-control combination varies in calibration success between stroke score and actual proportion of patients at each probability.

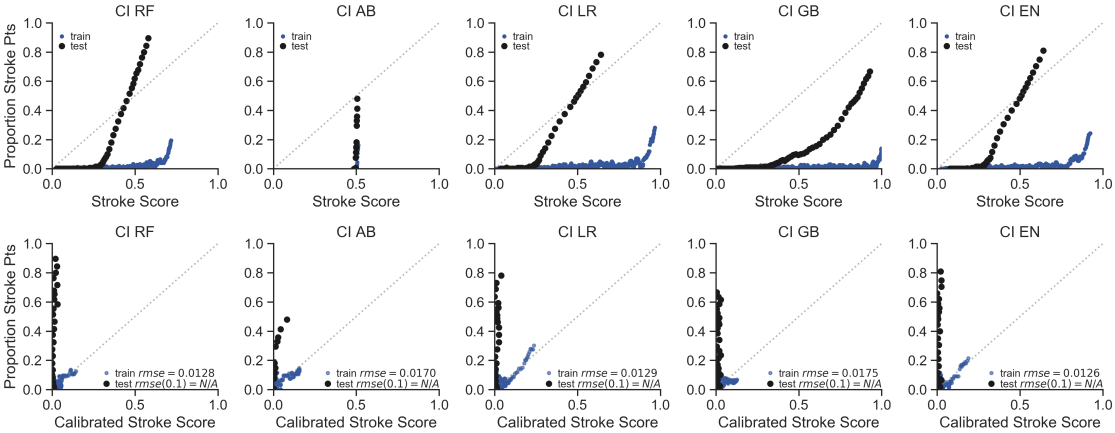


Fig. 19: Classifier type with CI case-control combination varies in calibration success between stroke score and actual proportion of patients at each probability.

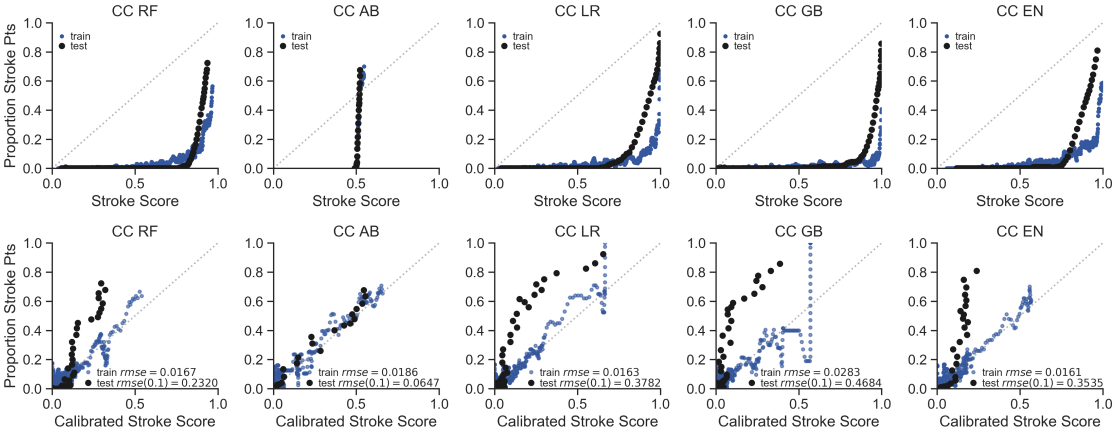


Fig. 20: Classifier type with CC case-control combination varies in calibration success between stroke score and actual proportion of patients at each probability.

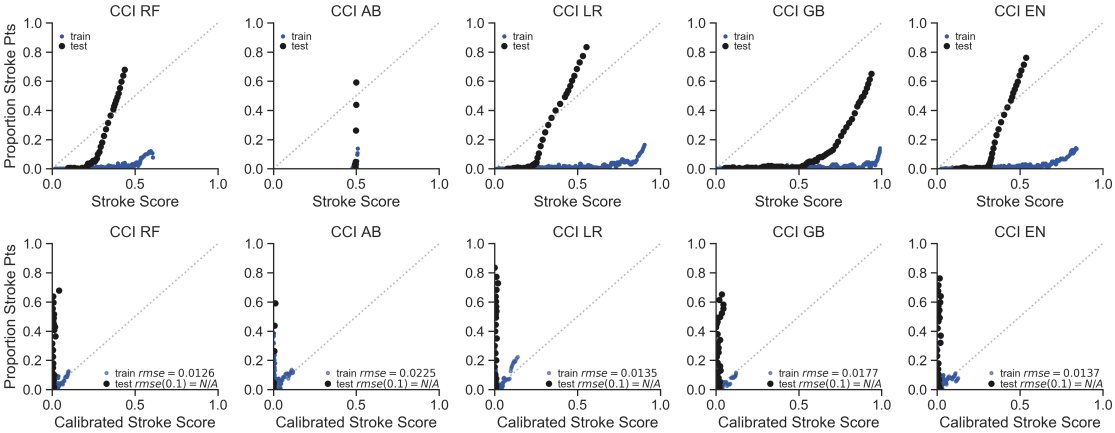


Fig. 21: Classifier type with CCI case-control combination varies in calibration success between stroke score and actual proportion of patients at each probability.

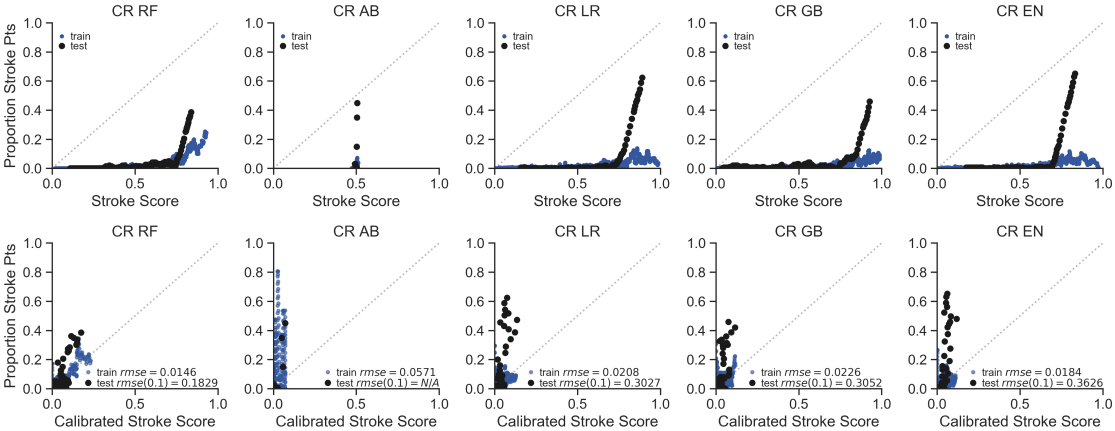


Fig. 22: Classifier type with CR case-control combination varies in calibration success between stroke score and actual proportion of patients at each probability.

Case/ Control Combo	LR Precision @50	RF Precision @50	AB Precision @50	GB Precision @50	EN Precision @50	LR Precision @100	RF Precision @100	AB Precision @100	GB Precision @100	EN Precision @100	LR Precision @500	RF Precision @500	AB Precision @500	GB Precision @500	EN Precision @500	LR Precision @NCases	RF Precision @NCases	AB Precision @Ncases	GB Precision @Ncases	EN Precision @Ncases
SN	1	1	1	1	1	1	1	1	1	1	1	0.91	0.98	0.98	0.98	0.84	0.72	0.83	0.82	0.85
SI	1	1	1	1	1	1	1	0.99	1	1	1	0.97	0.98	1	0.99	0.89	0.8	0.88	0.89	0.88
SC	1	1	1	1	1	1	1	1	1	1	0.99	0.96	1	1	0.99	0.89	0.79	0.9	0.9	0.87
SCI	1	1	1	0.88	1	1	1	1	0.89	1	1	1	1	0.88	1	0.92	0.9	0.91	0.8	0.91
SR	1	1	1	0.98	1	1	1	1	0.99	1	1	0.99	1	1	1	0.92	0.85	0.91	0.89	0.92
TN	1	0.09	0.63	0.78	1	1	0.11	0.66	0.76	1	0.95	0.13	0.86	0.57	0.9	0.69	0.14	0.75	0.45	0.67
TI	1	1	1	0.98	1	1	1	1	0.96	1	0.99	0.97	0.96	0.91	0.99	0.86	0.83	0.81	0.79	0.87
TC	1	1	1	0.96	1	1	1	1	0.95	1	0.99	0.98	0.99	0.9	0.99	0.87	0.85	0.87	0.77	0.88
TCI	1	0.96	0.48	0.95	1	1	0.97	0.44	0.92	1	0.97	0.88	0.48	0.74	0.98	0.72	0.7	0.38	0.56	0.74
TR	1	0.97	0.98	1	1	1	0.94	0.98	1	1	0.96	0.64	0.92	0.92	0.98	0.71	0.45	0.66	0.67	0.72
CN	1	0.03	0.52	0.46	0.99	1	0.03	0.6	0.41	0.96	0.94	0.03	0.67	0.35	0.75	0.72	0.03	0.6	0.31	0.53
CI	1	0.99	1	0.92	1	1	0.99	1	0.88	1	0.99	0.98	0.75	0.76	1	0.82	0.85	0.5	0.62	0.84
CC	1	1	1	0.95	1	1	0.99	1	0.96	1	0.99	0.89	0.99	0.91	0.99	0.84	0.74	0.85	0.76	0.86
CCI	1	0.98	0.9	0.96	1	1	0.99	0.77	0.9	1	0.99	0.95	0.83	0.7	0.99	0.79	0.75	0.7	0.55	0.75
CR	1	0.89	0.53	0.9	1	1	0.82	0.6	0.85	1	0.87	0.49	0.72	0.58	0.9	0.66	0.33	0.53	0.42	0.67

Table 7: Precision at top 50, 100, and number of known cases for each classifier. Blue values > 0.99, yellow values between 0.9 and 0.99, orange values between 0.7 and 0.9, red values < 0.7.